THE UNIVERSITY
of EDINBURGH

Genetic responses to environmental stress underlying major depressive disorder

Aleix Arnau Soler

THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I hereby declare that this thesis has been composed by myself and that the work presented within has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work are indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

I composed this thesis under guidance of Dr. Pippa Thomson. Chapter 2 has been published in *PLOS ONE* and is attached in the Appendix A, chapter 4 and chapter 5 are published in *Translational Psychiatry* and are attached in the Appendix C and D, and I expect to submit chapter 6 as a manuscript for publication. Therefore, the use of "we" is retained in these chapters. Only minor changes have been applied in comparison to the submitted manuscripts to keep a uniform format along the entire body of the thesis. Co-authors to these studies are indicated below. Co-authors contributed with the collection, access and preliminary curation of some raw data and/or providing critical revisions. Dr. Pippa Thomson also contributed to the initial draft of the introduction in chapter 2 and with supervision of all chapters. I confirm that I performed all the analyses and wrote the draft manuscripts myself.

Dr. Pippa Thomson, Dr. Caroline Hayward (my 2nd supervisor) and Dr. Mark Adams co-author chapters 2, 4, 5 and 6. Prof. Andrew McIntosh co-author chapters 4, 5 and 6 and was the Chair of my thesis committee. Dr. Toni-Kim Clarke, Dr. Donald MacIntyre, Mr. Keith Milburn and Dr. Lauren Navrady co-author chapters 4 and 5. MSc Erin Macdonald-Dunlop co-author chapter 5. Generation Scotland and the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium co-author chapters 2, 4, 5 and 6.

Aleix Arnau Soler 10th January 2019

Abstract

Major depressive disorder (MDD) is a common psychiatric disorder and a leading cause of disability worldwide. Such illness is the result of a complex interplay between genetic susceptibility and environmental risk factors. Adverse life events are experienced before the onset of depressive episodes in most patients, with robust evidence for the role of stressful life events (SLE) as a main trigger of depressive symptoms. However, not all individuals develop depression after episodes of stress. Thus, an individual's sensitivity to stress is an important predictor of stress response that may mediate the association between stress and depression. A deeper understanding of the genetic mechanisms underlying stress-sensitivity and stress response is, therefore, crucial to a better understanding of MDD and thus to improve treatments for both depressive symptoms and other stress-related conditions.

This PhD thesis uses empirical data from white Caucasian population-based samples. By incorporating in new hypothesis-free genome-wide association studies and polygenic approaches quantitative measures of recent SLE and neuroticism---a personality trait though to mediate or moderate the effects of adversity on depression risk---, this PhD thesis identifies the genetic influences to a proxy for sensitivity to environmental stress and genotype-by-environment interaction (GxE) effects underlying depressive symptoms.

Following an introductory chapter, **chapter 2** conceptualizes a proxy for our sensitivity to negative outcomes by modelling the interaction between genetic variants and MDD status on neuroticism score through a genome-wide interaction study. This chapter seeks to identify genetic variants contributing to a potential endophenotype mediating the associations between stress and depression, and examines whether genetic effects on such proxy for stress sensitivity partially explains the genetic contributions to liability not attributable to additive main effects. The strongest signals came from genetic variants associated with the glucocorticoid receptor function. Therefore,

Chapter 3 assesses the enrichment of the genetic contributions to liability of MDD within three glucocorticoid-related gene sets: one gene set reflecting "up-stream" cortisol signalling genes and two gene sets reflecting "down-stream" cortisol response genes**. Chapter 4** empirically tests and assesses the *diathesis-stress* theory for depression; using polygenic risk scores weighted by the additive effects of MDD derived from the Psychiatric Genetic Consortium MDD genome-wide association study and self-reported measures on recent SLE. This chapter provides evidence for the presence of GxE effects between stress and common genetic variants on risk of depressive symptoms. The empirical support for this theory validates other GxE approaches applying a genome-wide approach to investigate the causative effect of stress in the development of depressive symptoms. Thus, **chapter 5** presents findings from genome-wide by environment interaction studies in two cohorts that seek to identify common variants displaying an increased risk of liability to depressive symptoms in response to SLE. Whether inclusion of GxE effects improves the prediction of liability to MDD over that explained by genetic additive main effects alone is also tested. Furthermore, two potential forms of gene-environment interplay (i.e. GxE and gene-environment correlation) and their biological interpretation are extensively discussed. Stress contributes to many human conditions. Therefore, the GxE effects are also used to predict other stress-related physical and mental conditions. This chapter reports evidence of a potential shared aetiology between depression and other traits, such as schizotypal personality or heart disease, due to genetic mechanism underlying the effects of SLE. Finally, **chapter 6** brings back the *diathesis-stress* model investigated in **chapter 4**. This chapter incorporates into the diathesis framework the genetics effects for stress sensitivity and stress response estimated in **chapters 2** and **chapter 5**, respectively, and assess their relevance to the *diathesis-stress* theory. Genetic differences between women and men in stress response underlying the aetiology of depression are also discussed.

Genetics plays a significant role in the effects of stress. The findings presented in this thesis emphasize the relevance of genetic effects for stress sensitivity and stress response in depression and health in general. Overall, this thesis presents a range of original studies in order to advance our understanding of the genetic response to stress, comprehensively discussing the limitations and pitfalls of this research area, and provides a basis for future lines of research on gene-environment interplay in psychiatry.

Lay Summary

Depression is a common mental disorder that impacts on individuals, families and society. Genetic and environmental factors combine to increase the risk of suffering from depression. Indeed, before developing symptoms, most patients are affected by adverse or stressful environments. However, not all individuals that are exposed to adverse circumstances get depressed, suggesting differences in vulnerability to stress. One can think about this sensitivity towards the effects of stress as a behavioural or personality trait that mediates the association between stress and depression. Hence, several genes may influence such trait. A genetic component is behind the differences in how people respond to the effects of stress. Some genes modulate the effects of negative environments incrementing the risk of illness. This is known as gene-by-environment interaction effect. In this PhD thesis, the theory that genetic variation and stressful life events combine to increase the liability to depression is investigated. Multiple methods are explored in order to identify the nature of the genetic variants underlying both sensitivity and response to the effects of stress, as well as their importance in predicting depressive symptoms, in white European population. Genetic response to stress plays a key role on the onset and development of depression. Identifying genetic mechanisms underpinning the causative effects of stress may help to advance our understanding, not only of depression, but also of other disease of which stress is a main trigger, and thus, improve current treatments. This PhD thesis encompasses several original studies on this area of research seeking along the entire genome those genes contributing to the responses to environmental stress, particularly in depression.

Acknowledgement

First and foremost, I would like to express my gratitude to my supervisor, Pippa Thomson, for the opportunity she gave me to enrol this project, her guidance in the journey and her support and advice since the beginning.

Many people have helped and supported me over the past four years; too many to mention them all here, but they know who they are.

I am deeply grateful to the Institute of Genetics and Molecular Medicine of Edinburgh for funding my PhD and host me. I would like to thank all members and staff who helped me along these years and made this process easier and more enjoyable. I would like to thank the members of the Division of Psychiatry at the Royal Edinburgh Hospital, who contributed to this project.

None of this would have been possible without the time and effort of many anonymous participants who volunteered to take part in the cohorts used for my research. I thank them all and acknowledge all volunteers, technicians, clinicians, health care assistants, nurse and scientists, among others, who contributed to collect and make the data available.

This journey was shared with many new friends I made during my time in Edinburgh, within the institute and outside. I deeply thank all of them, including my old friends, some for helping me or stimulating me as we were going through the same process, others just for making me enjoy my time and being part of my life, past and future.

Last but not least, I want to thank my family for all their unconditional encouragement, love and unyielding support over these years. Specially, my mum, for her love and patience; my brother, for his inspiration and endless support; and my father, who saw me departing into this journey but unfortunately could never see the end. He would be proud. I love you all.

Publications

The following publications arose during my PhD training and form part of this thesis:

- Arnau-Soler, A. *et al.* Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS One* 13, e0209160 (2018).
- Arnau-Soler, A. *et al.* A validation of the diathesis-stress model for depression in Generation Scotland. *Translational Psychiatry* 9, 25 (2019).
- Arnau-Soler, A. *et al.* Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational Psychiatry* 9, 14 (2019).

The following article is expected to be submitted to *PLOS Genetics* and be deposited on *bioRxiv* during the following days.

- Arnau-Soler, A. *et al.* A new test of the diathesis-stress framework for depression: contributions to liability from genetics underlying sensitivity and response to psychosocial stress (2019).

During my PhD, I also contributed to the following articles:

- Hall, L.S. *et al.* Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational Psychiatry* 8, 9 (2018).
- Ryan, N.M. *et al.* DNA sequence-level analyses reveal potential phenotypic modifiers in a large family with psychiatric disorders. *Molecular Psychiatry* (2018).
- Clarke, T.K. *et al.* Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism. *Wellcome Open Res* **3**, 11 (2018).

Abbreviations

DNA	Deoxyribonucleic acid
DSLE	Dependent stressful life events
DSM-IV	The Diagnostic and Statistical Manual of Mental Disorders, fourth edition
DSM-IV-TR	The Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision
DSM-V	The Diagnostic and Statistical Manual of Mental Disorders, fifth edition
EPQ	Eysenck Personality Questionnaire
ERV	Endophenotype Ranking Value
ExE	environment-environment interaction
GCTA	Genome-wide Complex Trait analysis
GHQ	General Health Questionnaire
GREML	genomic restricted maximum likelihood
GS	Generation Scotland
GS:SFHS	Generation Scotland: Scottish Family Health Study
GWAS	Genome-wide association study
GWEIS	genome-wide by environment interaction study
GWIS	genome-wide interaction study

GxE	gene-environment interaction; also genotype-by-environment interaction
GxExE	gene-environment-environment interaction
GxG	genotype-by-genotype interaction
GxGxE	genotype-by-genotype-by-environment interaction
h^2	additive or narrow-sense heritability
h^2_{SNP}	SNP-based heritability (narrow-sense)
H^2	broad-sense heritability
HPA	the hypothalamus-pituitary-adrenal (axis)
ICD	International Classification of Diseases
ISLE	Independent stressful life events
LD	linkage disequilibrium
MDD	major depressive disorder
MDE	major depressive episode
OMIM	Online Mendelian Inheritance in Man
PGC	Psychiatric Genomics Consortium
PHQ	Patient Health Questionnaire
PRS	polygenic risk score
PRS_D	polygenic risk score weighted by depression score effects
PRS_{joint}	polygenic risk score weighted by joint effects (i.e. combined additive and GxE effects)

PRS_{MDD}	polygenic risk score weighted by MDD effects derived from PGC
PRS_{N}	polygenic risk score weighted by neuroticism effects
PRS_{SS}	polygenic risk score weighted by stress-sensitivity effects
PRSxE	polygenic risk score by environment interaction
PRSxSLE	polygenic risk score by stressful life events
r_g	genetic correlation
s.e.	Standard error
SLE	Stressful life events
SNP	single-nucleotide polymorphism
TSLE	Total stressful life events
UKB	UK Biobank

Contents

List of Tables

List of Figures

Chapter 1 Background

1.1 Introduction to major depressive disorder

Major depressive disorder (MDD; OMIM 608516), sometimes called major depression, clinical depression or unipolar depression, is a common psychiatric condition that debilitates and affects patient's mental health. In 2017, the World Health Organization ranked depression as the single largest contributor to global disability[1]. This, included the associated suffering to both patients and relatives, the deteriorated function and quality of life[2], medical morbidities[3,4], the potential mortality[5,6], and the global economic impact of the illness[7,8], and identifies depression (and mental health in general) as a public-health priority problem that must be faced and cannot be ignored.

MDD is a complex disorder with an aetiology resulting from multiple (often common and low-penetrance) susceptibility loci in combination with many environmental risk factors that may act and interact at multiple levels (i.e. genetic, physiological, psychological, social, cultural)[9,10]. It is well recognised that both genetic and environmental risk factors play a critical role in the development of MDD, and mental illness in general[10-14]. To fully understand the aetiological mechanisms underlying MDD we must integrate insights from multiple approaches and address the aetiology of MDD from different explanatory perspectives. This includes not only the study of genetic and environmental risk factors, but the integration of their interplay[15]. Therefore, investigating genetic responses and gene-environment interactions (GxE) between susceptibility loci and environmental risk factors, as in this thesis, is part of approaches aiming to provide insights into the aetiology of MDD.

1.1.1 Clinical features of major depressive disorders: symptoms and diagnosis

According to *The Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision* (DSM-IV-TR)[16], the manual most widely used by mental health professionals for diagnosing clinical MDD based on symptoms criteria, MDD is defined by a single or recurrent major depressive episode (MDE) without a history of manic, hypomanic, or mixed episodes. The essential feature of a MDE is the appearance of either depressed mood or anhedonia (or both) in nearly all tasks and situations during the day, nearly every day, and over a period of at least two consecutive weeks. In children and adolescents, however, such mood may be exhibited as irritability rather than low or empty mood. In addition, to be diagnosed with a MDE, the affected patient must present nearly every day at least four (or three if the patient presents both depressed mood and anhedonia) of seven additional symptoms drawn from a list (see **Table 1.1**). All these symptoms must either be newly exhibited or must be clearly exacerbated compared with the patient's pre-episode condition. They cannot be included if they are better account for by bereavement, or due to the direct psychological effect of a medical illness or substance abuse such as illegal drugs, alcohol or medication. The overall symptomatological panel must coexist with clinically significant distress or impairment in daily functioning (i.e. personal, familial, social, occupational or others) to the patient. Depending on the number of episodes, MDD can be classified as *single episode* or *recurrent* MDD if there are at least two MDE manifested over a lifetime. Recurrent episodes must be separated by a period of two consecutive months without meeting criteria for a MDE to be considered as an independent episode. MDD includes a broad range of illness, from mild to moderate or severe depression, depending on the number and severity of MDE. The precise criteria according to DSM-IV for a MDE are provided in **Table 1.1**.

Table 1.1 DSM-IV criteria for a major depressive episode (MDE)[16]

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure. Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.
Symptoms:

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad or empty) or observation made by others (e.g. appears tearful). Note: In children and adolescents, can be irritable mood
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)
3. Significant weight loss when not dieting or weight gain (e.g. a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. In children, consider failure to make expected weight gains
4. Insomnia or hypersomnia nearly every day
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
6. Fatigue or loss of energy nearly every day
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms do not meet criteria for a Mixed Episode
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning
D. The symptoms are not attributable to the physiological effects of a substance or a general medical condition
E. The symptoms are not better accounted for by bereavement

In DSM-IV, MDD is classified into the category “depressive disorders” along with *dysthymia*, a chronic and persistent mild depressive disorder (less intense and more persistent than MDD), and *depressive disorder not otherwise specified.* Depressive disorders are characterised by an overwhelming feeling of sadness, empty or irritable mood, and loss of interest

or pleasure in usual activities severe and persistent enough to interfere on daily functions, often accompanied by feelings of guilt, low self-esteem, tiredness, sleep or appetite disrupted patterns and cognitive impairment. Briefly, *dysthymia* consists of a milder but more persistent type of depressive disorder than MDD. It is defined by the presence of depressed mood nearly all day, for more days than not, during a period of at least two years (1 year for children or adolescents), along with two (or more) symptoms from a list including: diminished appetite or overeating; insomnia or hypersomnia; fatigue or low energy; reduced self-esteem; poor concentration or indecisiveness; and feelings of hopelessness. During the two year time period, the patient must never be without the manifested symptoms for more than two months at a time. *Depressive disorder not otherwise specified* (called "unspecified depressive disorder" in DSM-V[17]) encompasses any depressive disorder that fail to meet criteria for any other depressive disorder as, for example, in those patients that meet fewer than the required five symptoms for a diagnosis of MDD for two or more weeks. At the same time, in DSM-IV, depressive disorders are classified along with bipolar disorders into the diagnostic term "mood disorders". Depressive disorders are distinguished from bipolar disorder by the lack of an episode of mania or hypomania. Unlike DSM-IV, DSM-V classifies "depressive disorders" into a unique section adding, along with *MDD* and *dysthymia*, new depressive disorders classified by specific symptoms criteria (i.e. *disruptive mood dysregulation disorder*, *other specified depressive disorder* and *unspecified depressive disorder*), as well as by aetiology (i.e. *premenstrual dysphoric disorder*, *substance/medication-induced depressive disorder*, or *depressive disorder due to another medical condition*).

As we can see, the diagnosis of MDD is not easy or straightforward, and indeed, protocols and criteria vary across countries. Other widely used diagnostic tools such as *The International Classification of Diseases* (ICD)[18] maintained by the World Health Organization, are also applied to diagnose MDD and mental disorders in general. Completely opposite symptomatological panels (e.g. psychomotor retardation, weight gain and

hypersomnia; psychomotor agitation, weigh loss and insomnia) can result in the same “diagnosis of MDD” for patients who may need different treatments[19]. In fact, there are at least 227 possible combinations to meet DSM-IV (or DSM-V) criteria for MDE, many more if one takes into account subcategories of each symptom (i.e. severity, persistence, recurrence)[20]. Indeed, the current diagnosis of MDD may be too broad, resulting in significant clinical heterogeneity. Therefore, MDD cannot be seen as a clinical homogeneous entity, an issue widely discussed and investigated nowadays[21]. The range of symptoms I have just described with varying degrees of severity from severe to milder forms of MDE lying along a continuum, often accompanied by a range of comorbidities, suggest to some extent the existence of a depressive disorder spectrum of which the boundary with other depressive and mood disorders is confused[22]. Moreover, evidence suggests that psychiatric disorders lie on a continuous and broad spectrum of overlapping mental illness where the causal genetic influences transcend clinical diagnostic boundaries defined by DSM-IV/V and ICD-10. Consequently, there is a growing debate about moving from a symptoms-based nosology towards an aetiology-based nosology for psychiatric disorders, including MDD[23].

MDD is a treatable condition, although often it goes unreported and untreated. Symptoms can respond to pharmacotherapy or psychotherapy, with greater effectiveness when a combined treatment is applied[24]. However, the effectiveness of antidepressant treatments is imprecise. Almost all treatments are entirely symptomatic-based and focus on treating the symptoms rather than the aetiology. Indeed, recent studies suggest that taking antidepressants alone is ineffective in up to 40% of patients[25]. Whereas some suggest that the benefits appear to be greater the more severe the MDE is (i.e. treatment may be effective against chronic and severe depression, but probably fail on treating mild depressive episodes)[26,27], others disagree and suggest that effectiveness is not dependent on severity[28]. Either way, the adequacy and quality of antidepressant treatments need to improve[29]. Improved preventions and

more effective treatments targeting and treating the underlying pathophysiological cause, rather than focussing on relieving the symptoms perceived, are required. Therefore, insights derived from this thesis may provide better understanding of the aetiology of MDD, which eventually permit the application of better treatments and targeted provision of prevention strategies, at least, in a subtype of MDD patients.

1.1.2 Epidemiology and cost of major depressive disorder

Recently, the World Health Organization has estimated the number of people living with depressive disorders as 322 million[1], 3.5% more than estimates from 2015[30]. The number of individuals suffering depressive disorders is likely to keep increasing in the near future, with many people not only suffering from depression but also from other stress-related conditions simultaneously such as anxiety disorders, thus becoming the leading cause of disability worldwide, particularly in countries with low and middle income[31].

MDD is the commonest psychiatric illness worldwide, with lifetime prevalence estimated to be about 14%; although it varies widely across populations, countries and with socioeconomic status: from 1% in Czech Republic to 16.9% in US, with a 12-month prevalence ranging from 0.3% to 10%, respectively[9]. Overall, the global point prevalence of MDD has been estimated at 4.7% (4.4–5.0%)[32], with its highest point in older adulthood. Studies looking at prevalence of MDE estimated the average 12-month prevalence at 3.2% in patients without associated comorbidities, but it increased to 23% in patients with chronic comorbidities[4]. In general, the prevalence seems to be higher in high-income countries, although no difference is found in 12-month prevalence estimates, which may reflect higher persistence of MDE in low- and middle-income countries. Prevalence is also higher in people living in conflict areas[33]. However, this wide variability is likely due, in part, to a combination of factors linked to the epidemiological survey design (e.g. diagnostic criteria used, measurement applied, and sample selection among others)[32].

MDD can affect people of all ages, including preschool-aged children with minimum age of onset manifesting as early as 3 years[34]. However, the average age of onset is between 25-34 years[33,35]. Patients with earlier age of onset often show more severe symptoms and experience more lifetime episodes, among other features (e.g. higher levels of neuroticism, anxiety, paranoid or compulsive behaviours), than patients with later onset[36-38]. Early-onset is associated with increased social and occupational impairment, poor life quality, decreased self-esteem, greater suicidal ideations, more suicide attempts, and greater medical and psychiatric comorbidity, which overall results in greater health impact[39]; as well as increased risk in first-degree relatives and higher heritability.[40-42] The global distribution of age of onset of first MDE suggest heterogeneity by age of onset in MDD, showing plausible earlier-childhood-onset and late-adult-onset subtypes[43,44]. The former, has greater genetic overlap with more severe psychiatric disorders such as bipolar disorder and schizophrenia[45], as well as higher co-morbidity with personality disorders; whereas late-adult-onset is characterized by higher prevalence of preceding stressful life events (SLE)[46].

Gender-specific differences between women and men are also present in many aspects of the disorder, from manifestation of symptoms and course of illness to treatment response[47,48]. Women show a prevalence of MDD around 2-fold times greater than men, they are at higher risk (approximately 1.5- 2-fold increased risk compared to men), often report more symptoms, and have different coping skills (e.g. whereas women tend to cope through verbal and emotional strategies, men tend to cope by doing sport and consuming alcohol)[49-54]. In addition, MDD is associated with pregnancy[55]. MDD with postpartum onset (postpartum or postnatal depression; according to DSM-IV symptoms must appear in the first 4-6 weeks) is estimated to affect 10-15% of women[56,57], although it is not limited to mothers and also affects fathers[58].

Psychiatric comorbid disorders are common in patients with depressive disorders[3,59,60]. Most patients with MDD suffer comorbid disorders such as anxiety disorders[59,61,62], substance use disorders[59,63] and other depressive

disorders such as dysthymia[64]. For example, reports show that around 59% of individuals with lifetime MDD will suffer at least one episode of anxiety disorder (mostly generalised anxiety disorder or panic disorder), and 24% to 30% of cases will co-occur with substance use and impulse control disorders, respectively[59]. However, the list of possible psychiatric comorbidities is large, including (but not exclusive to): other anxiety disorders (i.e. agoraphobia, social phobia, obsessive-compulsive disorder, post-traumatic stress disorder), alcohol dependence, psychotic disorders, antisocial personality, insomnia and eating disorders[3,65]. Comorbidity is also frequent with somatic diseases and conditions such as cardiovascular diseases[66], diabetes,[67] chronic pain[68] and inflamation[69]. Co-occurring conditions may contribute as risk factor to each other (e.g. MDD can both be triggered by inflammatory mechanisms, as well as trigger inflammatory processes[69-71]), which offers an opportunity to use those risk factors to stratify MDD based on associated risk factors, and at same time, underpin the pathophysiological mechanism underlying specific subtypes of MDD. It has been proposed that inflammation resulting from the hyperactivity of the hypothalamus-pituitary-adrenal (HPA) axis, as response to stressful stimulus, and adult depression may be part of the same pathophysiological process[72]. However, the directional relationship between co-occurring conditions is unknown. It may be that somatic illness causes chronic and intense pain and stress that trigger an MDE, and maintain it over time, or that the biological mechanisms underlying the somatic condition eventually cause such MDE (e.g. endocrine diseases such as Cushing or Addison syndromes where a dysfunction in stress-response mechanisms, or the decreased levels of dopamine in Parkinson's disease[73], may eventually lead to depressive symptoms), or both. Therefore, if comorbidities are untreated, the cost of MDD may increase in both direct and indirect ways[8].

1.2 The aetiology of major depressive disorder

There are still many questions about the aetiology of MDD that remain unresolved. It is well known that both genetic and environmental factors play a key role in the development of psychiatric disorders. More than 30 years ago, Kendler and Eaves already argued in their article *Models for the joint effect of genotype and environment on liability to psychiatric illness*[74] published in 1986 about the importance of both genetic and environmental influences, and the relevance of investigating how both come together to fully understand the aetiology of most psychiatric disorders, including MDD.

"It is our conviction that a complete understanding of the etiology of most psychiatric disorders will require an understanding of the relevant genetic risk factors, the relevant environmental risk factors, and the ways in which these two risk factors interact. Such understanding will only arise from research in which the important environmental variables are measured in a genetically informative design. Such research will require a synthesis of research traditions within psychiatry that have often been at odds with one another in the past. This interaction between the research tradition that has focused on the genetic etiology of psychiatric illness and that which has emphasized environmental causation will undoubtedly be to the benefit of both." (Kendler and Eaves, 1986)[74].

However, few decades later, how exactly environmental factors such as psychological stress interact on the molecular level with our genome in order to shape risk, and resilience to, MDD, which genetic mechanisms are involved in such interactions, or why under the same environmental pressure some people develop depressive symptoms while others do not, are just some of the questions about MDD that still remain elusive and psychiatric research should solve.

1.2.1 Epidemiological risk factors

While the aetiology of MDD is undoubtedly influenced by genetic factors, as I further discuss in the following section 1.2.2, there are many other factors (e.g. physical, psychological or environmental) that also increase liability to MDD. Family history of psychiatric disorders, as well as factors within the

family environment (e.g. parental depression, disturbed parent-child relationship, changes of family structure, violence or neglect), can contribute to depressive disorders, mostly in adolescence[75-78]. Factors related to an individual's physical or mental health (e.g. chronic pain, sleep disorders, fatigue or history of previous MDE) also increase the risk of depressive disorders. As said in the previous section, there are many comorbid disorders associated with MDD, so that physical and mental illnesses (e.g. anxiety, hormonal disorders, Parkinson's disease, cancer, stroke and heart attack, among others) can increase such risk. Several psychological and cognitive factors that alter judgment and perception (e.g. distorted perception of life experience and others' views, chronic low self-esteem, pessimism or incapacity to recognize personal accomplishment) also add greater liability to depressive disorders[78]. Similarly predisposing personality traits such as high neuroticism, borderline or avoidant personality disorders increase risk[79,80]. Some of these traits can be at same time conceptualized as endophenotypes for MDD, as discussed in further detail in section 1.4. These risk factors, including issues caused by substance abuse or misuse, being in a low socioeconomic status, preterm birth and low birth weight, among others[81], are influenced by their environment.

Twin studies indicate an environmental component to MDD[82]. Risk factors confirmed as individual-specific environmental risk factors for MDD include: little or no social support and a wide range of adverse and SLE such as childhood abuse or maltreatment, being exposed to traumatic events, premature parental loss, the loss of a loved one, relationship break-up, divorce, sexual abuse, financial difficulties, social issues, poverty and unemployment[75,78,79,83-90]. It has been determined that 63% of the variance in liability to MDD could be accounted by "environmental" influences specific to an individual (i.e. individual-specific non-genetic factors) that are not shared within families[91]. However, this estimate could be inflated due to measurement error at the expense of shared environmental or additive genetic influences[91].

1.2.1.1 Environmental stress

Most individual-specific environmental risk factors produce some amount of psychological stress to the individual who suffers them. By psychological stress, I refer to a state characterized by strong negative emotions evoking distress (e.g. anxiety, rage, fear, anger, etc.) and usually accompanied by physiological changes derived from environmental adversity (e.g. SLE, or childhood trauma). The role of stressful environments and the physiology of stress response systems (e.g. the HPA the axis) have been always closely related to MDD, often categorized within stress-related disorders[92]. SLE play an important role in the aetiology of MDD. SLE are consistently associated with many aspects of the illness, from the onset of MDE to remission and relapse, as well as the severity of each episode and the number of depressive symptoms manifested in both adults and adolescents[84,90,93-95]. For example, individuals reporting a SLE have been estimated to have odds ratio of 5.64 for the onset of MDD during the month following the event[93]. However, the final individual's measure of SLE applied in research must be interpreted with caution as its construction, which most of times is done through self-reported questionnaires, may be genetically influenced (e.g. through genetic control of exposure to stressful environments or genetic control of how the events are reported) and therefore, not representative of pure random environmental effect (i.e. self-reporting bias). Self-reported SLE, as well as how we perceive the associated stress, have been shown to be moderately heritable[96-98]. One approach to address this limitation is to divide SLE into those likely to be influenced by genetic contributions and those likely to be random or out of individual control[99,100]. Thus, two categories for SLE have been proposed: *dependent* SLE and *independent* SLE. Such categorization is based on whether experiencing the event could be the response to our own behaviour and symptoms (likely driven by genetic influences), in which case we could be playing and active role in experiencing such *dependent* SLE; or not, so the *independent* SLE experienced is out of our control and not genetically influenced[99]. However, it is possible that some genetic variation implicated in MDD may increase exposure (or reporting) to

stressful environments (both *dependent* and *independent*)[101,102]. Thus, individuals with higher risk for MDD may expose themselves into riskier and more stressful environments, or may have greater vulnerability to their depressive effects[103]. At same time, whereas *dependent* SLE have been proved to have a causal effect on liability to MDD[104], individuals with diagnosis of depressive disorders also report more *dependent* SLE[105-107].

Heritability of reported number of SLE has been estimated from twin studies at between 20 to 40%[108,109], with an average heritability estimated from 6 studies of 28%[97]. Heritability estimates are higher for *dependent* life events (43-45%; similar to those for depressive disorders) than for *independent* life events (7-18%)[109-111]. Couple and familial environments tend to have a higher influence on *independent* life events[110]. A recent study, of which I am co-author, showed that SLE are positively associated with MDD, both for self-reported *dependent* and *independent* SLE, with relative risk of MDD for those experiencing any SLE of 1.44, and rising up to 1.91 when reporting 4 SLE, compared to non-exposed individuals[98]. Such association was already shown decades ago using the equivalent terms “personal” and “network” SLE, instead of *dependent* and *independent* SLE[102]. More recently, SNP-based heritability (h^2_{SNP}) of self-reported SLE, the variation on self-reported SLE attributed to common genetic variance, has been estimated at between 8-29%[112,113]. In a sample of more than 2,500 unrelated European individuals enriched by MDD cases h^2_{SNP} was estimated at 29% (s.e. 0.16)[112], and in a sample of 7,179 African American women h^2_{SNP} was estimated at 8% (s.e. 0.04). The latter also provided evidence of a strong genetic correlation (r_g = 0.95, p = 0.01) between the number of SLE reported and MDD[113]. Thus, highlighting the overlap of genetic variation contributing to both MDD and self-reported SLE. The differences in heritability estimates may be consequence of differences in genetic architecture, measures applied or familial and environmental influences, among others, across the different populations analysed.

Important for the rationale of this thesis, the effect of SLE on MDD may be mediated by genetics or gene-by-environment interactions (GxE)[11,114]. Twin studies have suggested a genetic contribution to sensitivity to the depressogenic effects of SLE on liability to depression[79,103,115]. It has been shown that the effect of SLE on those with a depressed co-twin was greater in monozygotic twins than in dizygotic twins[116]. Therefore, the incorporation of knowledge about environmental stress into molecular approaches should improve the ability to predict MDD. This is a main aim of this thesis. From now on, as environmental stress I refer to stress caused by environmental and psychosocial events with a psychological stress component, excluding other forms of environmental stress.

1.2.2 Genetic susceptibility factors: from the origins to the present

The first evidence that genetics may contribute towards the development of MDD, categorized into "manic-depressive illness" at that time, was in the 1920s[117]. Since then, many different approaches have been employed to investigate the role of genetic risk factors on susceptibility to MDD, from family and twin studies to more recent molecular analyses such as linkage and association approaches, which nowadays take advantage of whole genome sequencing technologies, in order to localize and identify such genetic risk factors.

1.2.2.1 Family and twin studies: MDD as heritable disorder

In the early decades of the twentieth century, studies suggested that what we know nowadays as MDD results from genetic influences. It was seen that depressive symptoms aggregate within families, seeding the idea that depressive disorders may be heritable[118]. Many years later, using the highest-quality family studies, a meta-analysis estimated that first-degree relatives of patients with MDD have an increased odds ratio of 2.84 of developing the illness[119], with similar estimates of 2.26 in a recent study[120]. However, families tend to share environments (as well as genetics).

Therefore, family studies alone cannot differentiate shared environmental influences from purely genetic influences. To get robust evidence of a genetic contribution on liability, twin studies were performed comparing depression concordance rates between monozygotic twin (identical) and dizygotic twin (non-identical) pairs[119,121], strengthening prior evidence and highlighting the existence of genetic factors underlying the aetiology of MDD. Finally, by meta-analysing primary twin studies, additive heritability (h^2; also called narrow-sense heritability)[122], which captures only the proportion of trait variation due to additive genetic effects, was estimated to range between 31-42% for MDD[119]. Meanwhile, others addressed the apparently sex specific effect, showing that MDD is substantially more heritable in women (40-42%) than men (29-30%), although most of the genetic risk seemed to be shared between sexes with the genetic correlation estimated at 0.50-0.65[121,123,124].

1.2.2.2 Molecular genetic studies: MDD and its genetic complexity

Heritability alone does not provide information about the genetic architecture of MDD, its complexity or its mode of action. Once it was shown that genetic features play a role in the aetiology of MDD, genetic studies including linkage and association studies were performed to map and identify genetic risk factors. Linkage studies attempted to identify the chromosome regions containing genetic risk factors and map the chromosomal location of the susceptibility loci involved based on the co-inheritance within families of such loci. The advantage of this kind of study was its ability to detect genetic influences without any prior knowledge of the pathophysiology of the illness under assessment. However, the first linkage studies in the 70s and 80s found no significant results, mainly due to the high polygenicity and low effect of the genetic variants contributing to liability to MDD (linkage studies are more appropriate to Mendelian disorders)[125-127]. Nevertheless, since 2003 when the first three genome-wide linkage analyses of MDD were performed, several genome-wide linkage analyses have reported statistically significant findings across various regions. However, they showed inconsistent results between them and reported different chromosomic regions[126,128-134]. Probably, the most robust evidence of a susceptibility loci linked to MDD was

found in two independent studies[133,134] showing a linkage to the chromosomal region 3p25-26, which contains among many others, a gene known to encode a protein for the metabotropic glutamate receptor 7 (*GRM7*).

Association approaches are generally classified in two main categories: candidate gene studies and, more recently, genome-wide association studies (GWAS). They are designed to test whether specific alleles are more common among patients with illness than among healthy individuals in a case-control design (e.g. MDD patients versus healthy controls), or whether such alleles are associated with variation on a quantitative trait (e.g. number of depressive symptoms). As the allelic spectrum underlying MDD is so broad[13,135,136], association approaches assess genetic variants on three major classes: *common variants,* including single-nucleotide polymorphisms (SNPs) with allele frequencies ≥1-5%, generally with allelic odds ratio < 1.2; *rare variants,* including SNPs with allele frequencies <1% with effect sizes ranging from small to large effect sizes; and *structural variants*, including copy number variants, insertion/deletions and translocations. From now on, all association studies will refer to studies using common variants unless specified.

1.2.2.2.1 Candidate gene studies in depression

Association studies of common variants were first limited to candidate gene studies. Candidate gene studies are based on prior evidence about the biological function of a gene, or relatively small number of genes, in the liability to MDD. These studies were based on prior insight on the pathophysiological mechanisms underlying MDD, or on regions thought to be involved with depression. Therefore, candidate gene studies were largely limited to testing polymorphisms from genes with a potential role on biological systems targeted by antidepressants (e.g. neurotransmitter, neuroendocrine or neuropeptide systems). Hence, many genetic variants in many candidate genes have been tested for association with susceptibility to MDD. A meta-analyses of candidate gene studies reported some evidence of several genes

associated with MDD in pathways including: serotonergic and dopaminergic systems, calcium signalling, neuroplasticity, cell binding, drug metabolism, developmental processes, cardiovascular functioning, cellular stress response, and other cellular regulatory pathways, within others[137]. However, most of the findings from candidate gene studies were not replicable. Given that our knowledge of the biology underlying MDD was (and still is) poor, the low effect sizes expected of common variants and the relatively small sample sizes used, the probability that any gene tested was detected as relevant was low, as result of underpowered studies, with a high likelihood of false-positive hits. These facts limit this type of strategy and are reflected in the number of non-replicated findings[13,138-140].

1.2.2.2.2 Genome-Wide Association Studies (GWAS)

Unlike hypothesis-driven candidate gene studies, GWAS allow scanning of variants along the whole genome to seek common risk variants linked to liability of MDD in a hypothesis-free approach without any *a priori* knowledge required about the underlying biology or regions involved[141]. This is possible due to the development of DNA technologies able to genotype hundreds of thousands of SNPs across the genome at an affordable cost (below $100 per sample). Due to linkage disequilibrium (LD; a non-random association of alleles), alleles from many nearby SNPs between recombination hotspots co-segregate in blocks (or haplotypes). Alleles from the same haplotype are non-randomly associated and co-inherited together. Thus, nearby SNPs are strongly correlated, allowing representative SNPs (called tag SNPs) to be informative for most of the other SNPs within the same haplotype. This high correlation across SNPs allows genotype arrays to cover most common genetic variation in a specific genomic region by genotyping a subset of its informative tag SNPs. This strategy attempts to cover most of an individual's genome-wide genetic variation just by genotyping a preselected subset of its total number of SNPs. In a GWAS, such large number of SNPs generates a large number of hypotheses and statistical comparisons to be tested. To control for false discoveries, and thus consider a SNP as genome-wide significant, a conventional type 1 error threshold is applied at p-value < 5 x

10^{-8}. This approximates to a Bonferroni correction *p*-value = 0.05 divided by the estimated effective number of independent statistical tests (i.e. 1 million tests)[142].

The first GWAS of MDD dates back to 2009, when the genomes from 1,738 individuals diagnosed with MDD were compared against the genomes from 1,802 healthy controls[143]. Since then, many GWAS have been performed. Most of them lead by the Psychiatric Genomics Consortium (PGC)[144], founded after it became obvious that larger sample sizes were required to gain enough power to detect common risk variants of small effect. Therefore, PGC obtained larger sample sizes by combining samples from around the world (i.e. impossible to reach by a single research group). Nevertheless, and unlike other psychiatric traits such schizophrenia[145], the success of such GWAS and meta-analyses in identifying statistically significant genome-wide SNPs associated with MDD was initially poor[13,146]. In 2013, the PGC published an MDD mega-analysis involving 9,240 MDD cases and 9,159 controls. However, this study failed to identify robust and replicable findings, concluding that common loci accounting for 0.5% or more of the phenotypic variance could be rejected with 90% power at that sample size[147]. Hence, they focused on obtaining larger sample sizes. Later, GWAS meta-analysis of depressive symptoms involving more than 34,500 individuals from 17 independent samples collected by the CHARGE Consortium failed to report any genome-wide significant loci, concluding that only sample sizes larger than 50,000 individuals would have enough power to detect common variants associated with depressive symptoms[148]. It was not until 2015 that the first genome-wide significant findings on MDD were published. In fact, the first study to report genome-wide significant findings was not conducted by the PGC, but by the CONVERGE Consortium. Following a different strategy, CONVERGE used low-coverage whole-genome sequencing and focused on reducing the phenotypic and ancestral heterogeneity. CONVERGE conducted a GWAS with only 5,303 selected Han Chinese women with a severe subtype of recurrent MDD (and 5,337 screened controls). As a result, they identified and replicated in an independent Chinese sample the first two

loci to reach genome-wide significance[149]. However, the associated loci may have population specific effects, as the genetic variants identified (in chromosome 10 near *SIRT1* and in *LHPP* genes) are much less common in European populations. These loci were not identified in larger meta-analysis conducted by PGC in European samples[150,151]. A year later, another meta-analysis of depressive symptoms, now involving 161,460 individuals, identified two further new replicable findings[152]. In addition, a study using self-reported MDD, based on self-reported diagnosis or treatment for depression, on 75,607 "affected" individuals and 231,747 "unaffected" controls identified 15 associated loci, some of which were implicated in GWAS of related psychiatric illnesses[153]. Later, a meta-analysis conducted by PGC involving 130,664 individuals diagnosed with MDD and 330,470 controls identified 44 loci at genome-wide significance (containing almost 600 SNPs meeting statistical genome-wide significance), of which 6 loci were shared with schizophrenia[150]. Finally, the largest and most recent meta-analysis to date based on 246,363 cases and 561,190 controls detected 102 independent common variants (87 of which replicated in a further independent sample of 474,574 cases and 1,032,579 controls), 269 genes, and 15 gene-sets associated with depression[151]. Despite of the success from the CONVERGE study using a refined and apparently less heterogeneous subset in detrimental of sample size, maximizing the number of individuals involved is likely to provide better insights, at least on the currently available sample sizes[154].

Together, these studies reflect how difficult has been to identify causal loci for MDD and how essential it is to collect larger sample sizes. As for most human complex traits[155,156], the polygenic architecture of MDD involves many genetic loci of small effect. However, the low success in identifying genome-wide significant loci compared to other psychiatric disorders such as schizophrenia, which required 3-5 times less affected cases in order to detect similar number of significant findings, is due to many factors resulting in a lack of power[157]. For example, MDD is substantially more prevalent and less heritable than schizophrenia (prevalence < 1%, H^2: 65-85%), so assuming

that both disorders share similar number of common causal variants, their effect sizes on MDD must be smaller and therefore more difficult to detect. In addition, MDD is likely to be more genetically heterogeneous (i.e. due to misdiagnosis, or individuals sharing the same diagnosis or symptoms due to different genetic aetiologies), what substantially reduce the power to detect any association[158,159]. Hopefully, a combination of different strategies, incorporating new DNA sequencing technologies (i.e. whole- and exome-genome sequencing), will provide, in a near future, insights on the effects of the full spectrum of genetic variation, including rare and copy-number variants, on the aetiology of MDD.

1.3 Exploiting GWAS statistics summary data

Although, until recently, there has been poor success in identifying single risk variants for MDD or depressive symptoms, GWAS summary statistics can be used in combination with bioinformatics tools to perform downstream analyses and thus gain further insights on the genetic aetiology underlying MDD. For example, association signals from each single marker summarized from GWAS data, even if not significant, can be aggregated and translated into sets of scores, genes, systems or networks in order to examine genetic effects beyond the association of a single locus. Next, I present some of the most relevant bioinformatics tools applied in this thesis with some findings on depression.

1.3.1 Pathway and gene-set analyses

The signals detected in GWAS can be aggregated into functionally related gene sets that aggregate in specific biological networks, or into ontologically related genes that share specific features or attributes (e.g. biological process, molecular function or cellular component). Looking at the enrichment of the aggregated signal in these sets of genes, gene-set analysis can provide insights about pathways and networks underlying biological mechanisms, functions and components involved in the pathogenesis of MDD. Several pathway analyses have implicated pathways and networks likely to be involved in the underlying biological mechanisms of MDD including: protein phosphatase type 2A regulatory activity, cell and cell-cell junction organization, apical junction assembly, regulation of histone modification[160], neurotransmitter and neuronal systems, immune system, inflammatory response[161], negative regulation of transcription and nucleic acid metabolism[162], among many others[150,151]. Furthermore, many of the pathways identified have been also implicated in other psychiatric disorders suggesting a shared aetiology across psychiatric disorders[163,164].

1.3.2 Polygenic risk scores

Another powerful tool that benefits from GWAS data is the construction of polygenic risk scores (PRS; also called genomic scores or polygenic profile scores). PRS are individual single measures of common genetic risk burden for a specific trait or condition derived from GWAS summary statistics. PRS aggregate the number of risk alleles (i.e. unweighted PRS) for many genetic variants that exceed a selected *p*-value threshold (e.g. *p*-value < 1×10^{-5} or *p*-value < 0.05). The relative contribution of each allele (i.e. their effect sizes: odds ratios or betas) derived from the GWAS can be used to weight these risk alleles. Thus, weighted PRS aggregate the number of risk alleles weighted by their effect sizes. Whereas unweighted PRS assume that all risk alleles have the same effect and therefore contribute equally to develop the disorder, weighted PRS take into account the importance and contribution of each allele. Therefore, PRS takes into account the aggregated genetic risk from a large number of common risk variants that given the lack of power do not reach genome-wide significance. From now on, I will always refer to weighted PRS when citing PRS.

PRS have potential research, medical and clinical applications as predictors of MDD or as predictors of other diseases, conditions (e.g. prognosis, severity of symptoms or treatment response) and cross-disorder effects of common risk variants, providing insights on the genetic correlation and overlap of the disorder[165-167]. PRS have shown that liability to MDD, and the number of depressive symptoms, is increased in carriers of high burden of common risk variants associated with MDD[83,147,150,168,169]. To date, the largest GWAS for MDD have identified common variants that account for up to 1.2 - 1.9% of the variance in liability to clinical MDD[150,151]. However, the accuracy and the amount of variance predicted by PRS is dependent on the number of variants aggregated, the phenotypic diagnose used and, specially, the sample size of the discovery sample (i.e. the sample where the GWAS summary statistics were generated)[170]. Therefore, it is expected that PRS prediction explanatory power will increase with larger sample sizes[167,170].

1.3.3 SNP-based heritability and genetic correlation

SNP-based heritability (h^2_{SNP}) and genetic correlation (r_g) are important population parameters that can be estimated using either individual-level genotype data or only GWAS summary statistics without requiring of individual's raw genotype data. h^2_{SNP} is the variance in liability due to common additive genetic variation (i.e. narrow-sense heritability; h^2) attributable to the common risk variants used in a GWAS[171,172]. Therefore, h^2_{SNP} is often seen as a lower bound for the total h^2. r_g is the additive genetic effect shared between MDD and other traits and diseases. Statistically, it is estimated as the covariance between two traits captured by all SNPs scaled by the square root of the product of the genetic variance for each trait. In this thesis, I use two methods to estimate both parameters: linkage disequilibrium (LD) score regression[173,174], which only requires GWAS summary statistics, and genomic restricted maximum likelihood (GREML), implemented in the GCTA software, that uses individual-level genotype data to construct a genomic relationship matrix[175-177]. LD score regression uses data on LD correlations between SNPs extracted from a reference panel of the ancestry population under study in order to calculate an LD score for each SNP as the sum of its LD correlations with other SNPs. Under polygenic architecture, SNPs with high LD are more likely to tag a causal SNP and therefore expected to have higher test statistics than SNPs with lower LD. With a single-trait LD score regression, the h^2_{SNP} explained by all common variants, used to infer the LD structure, can be estimated as a function of the regression coefficient of all SNP association test statistics on their LD scores[173,178]. Thus, LD score regression can estimate variance/covariance components by regressing association test statistics of the SNPs detected in GWAS on their estimated LD scores. Cross-trait LD score regression exploits the relationship expected between two traits in order to estimate the genetic covariance required to calculate r_g[173,174]. Conversely, GCTA exploits genomic relationship matrix constructed based on common genetic variants using individual-level genotype data from unrelated individuals in order to estimate similarities between cases and controls and thus, to estimate genetic variance and

covariance structures[171,175-177]. The genomic relationship matrix captures the genetic relatedness between individuals and is the key component in the GREML method[179,180]. It uses a linear mixed model to compare pairwise similarities between case pairs and control pairs to case-control pairs in order to estimate genetic variation and to detect the aggregated effects of common variants affecting both traits. While LD score regression is much less computational demanding in terms of memory and time than the GREML approach, LD score regression is less accurate and report larger standard errors for the variance components estimated[174,181]. Therefore, if possible, it is recommended to apply the GREML approach rather than LD score regression. LD score regression assumes that discovery samples of GWAS are drawn from the same reference population used to estimate the LD structure. However, differences in LD structure between samples (genetic heterogeneity) can be present and, consequently, bias LD score regression estimations[181].

The lifetime risk (or prevalence) in a population is required to estimate h^2_{SNP} on the liability scale. The most recent estimates of h^2_{SNP} on the liability scale suggested that common variants account for 8.7 - 8.9% of the heritability (s.e. 0.004 and 0.003; using a prevalence of 15% and 0.3%, respectively[150,151]); showing enrichment of h^2_{SNP} particularly at highly conserved regions in mammals, and at intronic and H3K4me1 regions across the genome implicated in regulatory activity[150,151]. Enrichment was not detected in exons, suggesting that common exonic variants may not have a large role in the aetiology of MDD. A previous study from PGC comparing h^2_{SNP} estimates on the liability scale across psychiatric disorders, including MDD (9,041cases and 9,381 controls), estimated MDD h^2_{SNP} at 0.21 (s.e. 0.021; 15% prevalence), similar to h^2_{SNP} estimates between 15-30% in the other psychiatric disorders[182]. SNP-based coheritabilities between MDD and other psychiatric disorders were estimated at 0.47 (s.e. 0.06) with bipolar disorder, 0.43 (s.e. 0.06) with schizophrenia, and 0.32 (s.e. 0.07) with attention-deficit hyperactivity disorder. However, no significant SNP-based

coheritability between MDD and autism spectrum was detected. Therefore, these study provided evidence of genetic contributions and a partial shared genetic overlap, reflecting pleiotropy, between psychiatric disorders[182]. Another study, performed on 25,571 participants from UK Biobank, estimated h^2_{SNP} for moderate and severe recurrent MDD at 0.195 (s.e. 0.03)[183]. Surprisingly, h^2_{SNP} of single MDE was estimated as 0, also on males and females, respectively[183]. It must be note that all the estimates from this study were adjusted by a variable reflecting negative experiences in the previous 2 years, which suggests that environmental factors are the major cause of single MDE.

There is a strong genetic correlation (r_g = 0.85) between clinical MDD diagnosed by clinicians and self-reported definitions of depression[184]. A study estimating r_g between depression and 23 other phenotypes using LD score regression and GWAS summary statistics reported significant r_g of MDD with height (r_g = -0.13; s.e. 0.05, p = 0.01), triglyceride levels (r_g = 0.18; s.e. 0.08, p = 0.03), and other psychiatric disorders: bipolar disorder (r_g = 0.48; s.e. 0.11, p = 6.5 x 10^{-6}) and schizophrenia (r_g = 0.51; s.e. 0.07, p = 1.32 x 10^{-11})[174]. This finding supported the r_g estimated between MDD and other major psychiatric disorders in a former study using raw genotype data in GCTA[182]. This former study estimated significant r_g between MDD and other psychiatric disorders: bipolar disorder (r_g = 0.47; s.e. 0.06, p = 1.5 x 10^{-14}), schizophrenia (r_g = 0.43; s.e. 0.06, p = 6.0 x 10^{-15}), and attention-deficit hyperactivity disorder (r_g = 0.32; s.e. 0.07, p = 6.8 x 10^{-6})[182]. Recently, the best-powered analyses of r_g between MDD and other traits estimated significant r_g with a wide range of psychiatric disorders, medical diseases and human traits[150,151]. Among all psychiatric disorders tested, r_g was significantly positive thus providing more evidence that the aetiology of MDD overlaps substantially with other psychiatric disorders. Therefore, it reinforces what is well establish nowadays, that some specific common variants are associated with a wide range of psychiatric and stress-related disorders[92,166] and confer vulnerability towards a wide range of physical and mental conditions[23].

1.4 The endophenotype concept

The link between genetic contributions and the final clinical symptoms of MDD is largely unknown. A strategy used in psychiatric research to identify causal mechanisms between the genetic aetiology and the final clinical manifestation is the use of endophenotypes. “Endophenotype” is a term used for biological or psychological entities (some times called “internal phenotypes”) that may lie between the causal genetic factors and the eventual symptoms of the disease[185]. It builds on the assumption that some entities are aetiologically less complex than others. To be considered an endophenotype, such entity or biomarker must fulfil a range of criteria: it must be heritable, associated with the illness in the population, be largely state-independent (i.e. it manifests in the individual regardless of being the illness active or not), both endophenotype and illness must co-segregate together within families, and it must be found at higher relative risk in unaffected family members than in general population[185]. Therefore, endophenotypes are measurable biomarkers phenotypically associated and genetically correlated with liability to disease due to, at least in part, shared underlying genetic influences[185,186]. Overall, the initial hope was that endophenotypes would be less genetically complex (i.e. involving fewer genetic variants) and more closely related to the underlying aetiology than the clinical symptoms or phenotype of interest. Consequently, endophenotypes should improve the power to detect genetic influences on liability to psychiatric diseases, and the aetiology of endophenotypes should be easier to understand. Maybe occupying an intermediate position, endophenotypes would be linked to the causal pathway between genetic determinants and the final diagnosable symptoms of disorders, with larger genetic effect sizes, and being potentially closer to the genetic variability level. Nevertheless, even if the endophenotype as a concept is easier to understand or define than the phenotype under study (e.g. cerebral cortex thickness vs. intelligence), their aetiology has not been proven to be less complex. Indeed, evidence suggests that the assumptions behind the concept and its benefits are not

met[187]. Hence, the challenges faced in the study of psychiatric disorders due to their aetiological complexity arise as well on the study of the aetiology of endophenotypes, thus raising more questions.

A wide variety of endophenotypes have been proposed for MDD (e.g. learning and memory impairments; reduced reward functioning; increased stress-sensitivity; REM sleep abnormalities; functional and structural brain abnormalities; dysfunctions in serotonergic, catecholaminergic, HPA axis, and corticotropin-releasing hormone systems; and intracellular signal transduction measures)[188]. However, their true credentials as endophenotype were often not assessed or are unmet (e.g. there is little or no evidence of co-segregation within families of depression for most of the putative endophenotypes listed above). However, in 2012, Glahn *et al*. developed an empirically-derived metric called Endophenotype Ranking Value (ERV) to rank and select optimal endophenotypes for mental illness based on heritability estimates of both the endophenotype, the disorder of interest, and their shared genetic correlation[189]. They applied the ERV to recurrent MDD in a large set of putative behavioural/neurocognitive endophenotypes and reported as the top-ranked endophenotypes a score to measure severity of depressive symptoms called Beck Depression Inventory[190] and the score assessing neuroticism derived from the Eysenck Personality Questionnaire (EPQ)[191]. More recently, a review of selected candidate endophenotypes for depression reported moderate to strong evidence that neuroticism is a true endophenotype of depression[192]. The criteria was also met for morning cortisol, cortisol awakening response, and frontal asymmetry of cortical electrical activity; although they lacked evidence of heritability and familial co-segregation with depression, probably due to the lack of family and twin studies with such a focus[192].

1.4.1 Depressive symptoms

Measures of depressive symptoms show substantial heritability estimates, ranging from 15% to 50%[148,152,193,194]. As endpohenotypes, scores based on depressive symptoms should improve the likelihood of identifying genetic

factors contributing to MDD[189]. MDD results from combinations of numbers and levels of severity of depressive symptoms; MDD status and depressive symptoms are highly associated and share mostly the same aetiology. MDD shows a genetic correlation with measures of depressive symptoms up to r_g = 0.98[150]. Similar risk factors are consistently associated across different types of depression, from brief episodes to recurrent manifestations, including subthreshold depressive disorders[195]. Both clinical diagnoses of MDD and depression diagnose defined based on self-reported symptoms are genetically strongly correlated (r_g = 0.85)[184]. These definitions of MDD status show estimates of genetic correlation with measures of depressive symptoms around r_g = 0.8[184]. There is a high genetic correlation between depressive symptoms and well-being, and between depressive symptoms and neuroticism[152]. However, there are several instruments to assess the severity and number of depressive symptoms, with high sensitivity to detect cases of MDD[196], including, among others, the Beck Depression Inventory[190] sated above to define a top candidate endophenotype of MDD, or the General Health Questionnaire[197] further and extensively discussed in **chapter 4** and **chapter 5**.

1.4.2 Neuroticism

Neuroticism is one of the most promising candidate endophenotypes of MDD[198]; and a key component in **chapter 2** and **chapter 3** to derive a proxy for sensitivity to stress. It is a personality trait characterized by emotional instability with predisposition to experience and report negative emotions often accompanied by low self-esteem and feelings of negative affect (e.g depressive, anxious or guilty feelings)[199] defined as a tendency to cope poorly with stress and to experience feelings of sadness, anxiety, anger, irritability, self-consciousness, worry, hostility and vulnerability[200]. The most common strategy to construct a measurable entity for an individual's level of neuroticism is through the use of a self-reported questionnaire. In this thesis, the neuroticism scores applied were assessed using 12-item questions from the "neuroticism-stability" domain of the EPQ, that has a reliability greater than 0.8[191,201]. A score on a range of 0-12 reflecting a neuroticism level is

constructed by adding up the number of "Yes" responses from the self-reported EPQ revised short-form. The 12-item questionnaire is provided in **Table 1.2**.

Neuroticism scores are able to consistently distinguish depressed and non-depressed individuals, although only at the group level[202]. Consistent with the criteria required to be considered an endophenotype, neuroticism is highly heritable, with heritability estimates up to 54% (being greater for women than men)[189,203-207]. Broad-sense heritability based on the meta-analysis of six cohorts was estimated at 48% with strong evidence of non-additive genetic influences[204]. Up to 15% of the phenotypic variance in neuroticism has been attributed to a non-additive genetic component[205]. There is no consistent evidence for a shared environmental component, suggesting that neuroticism is influenced by genetic (i.e. additive and non-additive) and unshared environmental factors[205,207,208]. Narrow-sense heritability (h^2) estimates range from 22 to 43%[189,203]. A twin study estimated that 55% of the genetic variance in depression was shared with neuroticism[209]. There is evidence that high neuroticism and depressive disorders co-segregate within families[210] and are strongly correlated at the phenotypic and genetic level[198,211,212], with estimates of r_g at 0.7 and 0.74 in MDD and recurrent MDD, respectively[150,151,189]. Neuroticism levels are higher in depressed patients compared to non-depressed individuals[202,213,214], and there is evidence that neuroticism levels differ in individuals with depression during the depressive state and before or after symptoms manifest[215]. In addition, a Mendelian randomisation approach has reported a putative causal effect of neuroticism on depression, and also a putative causal effect of depression on neuroticism[151]. Neuroticism shows a substantial stable component across lifespan[216]. However, some studies report evidence for change as well as stability throughout our lifetimes[217-220]. Whereas the stable component of neuroticism is strongly determined by genetics, change in neuroticism score is attributed to the effects of unshared environment[217]. Persistent change in neuroticism score has been shown in response to life events, including SLE[218-220]. Therefore, neuroticism has been used as a measure of stress-

sensitivity[221]. There is evidence that neuroticism score mediates or interacts with the effects of adverse life events on risk of depression[79,222,223]. Individuals with higher neuroticism levels show higher risk of depression as long-term contextual threat of SLE increase[79].

Table 1.2 Eysenck Personality Questionnaire-Revised Short Form Neuroticism scale

1.	Does your mood often go up and down?
2.	Do you ever feel "just miserable" for no reason?
3.	Are you an irritable person?
4.	Are your feelings easily hurt?
5.	Do you often feel "fed-up"?
6.	Would you call yourself a nervous person?
7.	Are you a worrier?
8.	Would you call yourself tense or "highly strung"?
9.	Do you worry too long after an embarrassing experience?
10.	Do you suffer from "nerves"?
11.	Do you often feel lonely?
12.	Are you often troubled by feelings of guilt?

Personality traits may influence how individuals experience and self-report SLE. For example, neuroticism is correlated with greater reporting of SLE, with a more severe perception of the corresponding impact among individuals with high levels of neuroticism[223,224]. Therefore, it is possible that neuroticism increases both sensitivity to and/or reporting of SLE amongst individuals with depression. Neuroticism has been positively associated with number of SLE reported, both for *dependent* and *independent* events[98].

1.5 Gene-environment interplay: theoretical models

There is clear evidence that MDD is influenced by both genetics and environmental stress[14]. However, how both components interplay is complex and poorly understood. There are multiple theoretical models proposed in the literature to explain how genetics and the environment interact and how GxE affects liability to MDD. Next, I present the most relevant models for this thesis: the first theoretical models presented in 1986 for *the joint effects of genes and environments on liability to psychiatric illness* by Kendler and Eaves[74], the *diathesis-stress* model[225], and the *differential susceptibility* model[226,227].

1.5.1 Models from Kendler & Eaves, 1986

Kendler and Eaves were convinced, more than 3 decades ago now, that the aetiology of psychiatric disorders, including MDD, involves the interaction between the effects of an individual's genes and the environment[74]. As previously shown at the beginning of section 1.2., Kendler and Eaves argued that to entirely understand the cause, or set of causes, of such disorders requires the comprehension of both the relevant genetic risk factors, the relevant environmental risk factors and the ways in which they interact. Therefore, they proposed three basic theoretical models to conceptualize fundamentally different forms in which genetics and environmental risk factors may jointly influence liability to psychiatric disorders. Representative illustrations of each model are shown in **Figure 1.1**.

1st model: *additive effects of genotype and environment*. This model conceptualizes the additive contribution of genetic and environmental risk factors to liability. It has been the base of most epidemiological research carried on MDD since then. This model, seen as the simplest one, considers genetics and environmental factors as completely independent entities; therefore, the final risk of MDD would be the mere additive combination of their corresponding completely independent effects (**Figure 1.1**; left).

***2nd model**: genetic control of sensitivity to the environment*. This model is based on the idea that genes (or genomic features) may control sensitivity to environmental effects, although it can also be conceptualized as the environmental control of gene expression. In other words, one entity (i.e. genetic or environmental) alters the protective or risky influence of the other. Under this assumption, not only environmental stressors but also their corresponding genetic responses play a key role on the pathogenesis of MDD. This is the fundamental premise behind theories such as the *diathesis-stress* theory or the *differential susceptibility* perspective (explained below), studies of GxE, and this thesis in general (**Figure 1.1**; centre).

***3rd model:** genetic control of exposure to environment*. This model suggests that there are genetic influences on how we expose ourselves to certain environments, so genes may alter the probability of being involved in adverse situations and stressful environments. This genetic control may act through personality and behaviour. It is known that certain genes influence certain behaviours and personality traits (e.g. neuroticism and extraversion). These traits are associated with higher exposure to risk-predisposing environments, thus, acting as mediators between genes and the environment[97] (**Figure 1.1**; right). As shown later in section 1.6.2, this is a case of active gene-environment correlation (rGE), but not GxE. To better understand the nature of the causal relationship between genes and the environment some studies focus on exploring the heritability of the environmental exposure, which is measured through self-reported questionnaires[96,109].

However, and as Kendler and Eaves already highlighted, the effect of the genetics-environment interplay on liability to MDD cannot be entirely explained by one of these three "basic" models, but likely by more complex combinations of all them[74]. For example, the "fan-shaped" GxE detailed later in section 1.6.1 result from combining the effects conceptualized in the 1st and 2nd models.

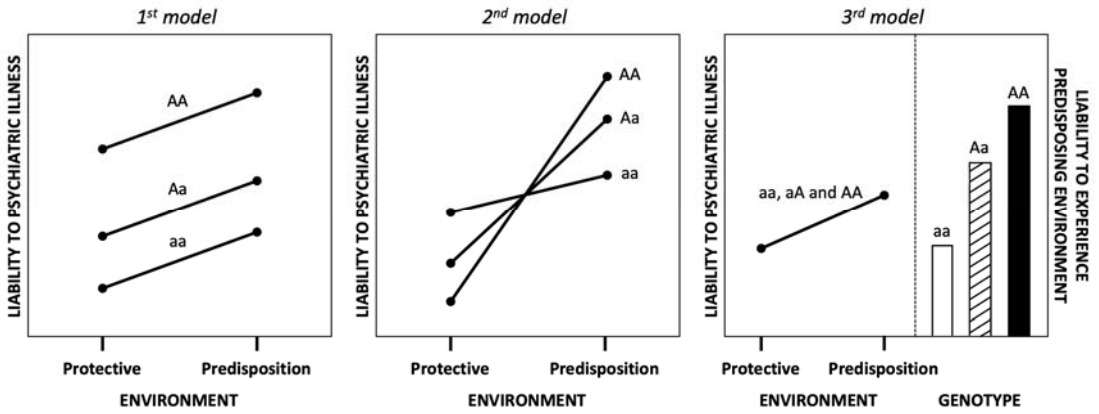


Figure 1.1 Theoretical models on liability to psychiatric illness from Kendler & Eaves, 1986. *1st model represents the liability to illness as a function of genotype and environment with additive effects of both the genotype and the environment (independent from each other). 2nd model represents the liability to illness as a function of genotype and environment with genetic control of sensitivity to the environment (alternatively, environmental control of gene expression). 3rd model illustrates the liability to illness as a function of genotype and environment with genetic control of mean liability to illness through genetic control of exposure to the environment. Figures are adapted from the original manuscript[74].*

1.5.2 The *diathesis-stress* model

Perhaps, the leading theoretical framework, and most widely investigated in GxE research, to explain the development of MDD is the *diathesis-stress* model. It goes back to 1960s, when it was first introduced in order to explain the development of schizophrenia[228]. The core basis of this theory is that genetic factors (i.e. an individual's inherent features) and environmental adversity (e.g. psychological stress) increase the liability to illness both independently and in combination; but it is only when combined that a liability threshold is reached and consequently the symptoms and illness manifest. Therefore, the predisposing effects of such adversity or stress would trigger the manifestation of symptoms only when combined with an inherent

vulnerability. *Diathesis* refers to this inherent vulnerability as a genetic vulnerability, predisposition or risk. In genetically predisposed individuals, their vulnerability to develop and manifest the disease will be increased when exposed to the effects of such environmental adversity, but it would not be enough to trigger the disorder in the absence of such environmental influences. On the other hand, individuals exposed to the same environmental exposure that do not develop the disease, either because they do not have this inherent diathesis or because they are under the influences of other protective factors, are resilient[225,228]. However, this model has received criticism for adopting a perspective where only the negative influences of the environment have an effect on liability and thus omitting the protective influences from positive environmental factors[226]. The *diathesis-stress* model is illustrated in red in **Figure 1.2**.

1.5.3 The *differential susceptibility* model

A more recent alternative theoretical framework is proposed by the *differential susceptibility* theory, so-called because instead of adopting a perspective of genetic vulnerability as the *diathesis-stress* theory, it adopts a perspective of a genetic susceptibility to the effects of both negative and positive environmental influences[226,227] (the *differential susceptibility* model is represented in **Figure 1.2**). In accordance with this theory, individuals who are more genetically susceptible to environmental influences would be more vulnerable to the negative effects of environmental adversity and consequently be at higher risk of developing MDD. However, these same individuals would also get higher benefits and higher protective effects from positive environmental influences (e.g. parenting or social support). Hence, the *differential susceptibility* theory supports genetic plasticity, as the presence of "plasticity alleles" that are more susceptible to both beneficial and detrimental effects from environmental influences, rather than "risk alleles"[229,230]. Recently, an individual's genetic response to positive environmental influences has also been conceptualize into a specific

theoretical framework called *vantage sensitivity*[231], in which only the positive influences of the environment have an effect on liability (see represented the *vantage sensitivity* model in blue in **Figure 1.2**). The benefits of such positive influences and protective effects on liability would vary as a function of their inherent features (e.g. genetic variants) and would be relevant for clinical intervention and treatment response[232].

As we see, the *differential susceptibility* theory takes into account the full spectrum of environmental influences from negative to positive exposures. Whereas the *diathesis-stres*s theoretical framework would fit the *differential-susceptibility* theory when you only consider negative environmental exposures, the *vantage sensitivity* theoretical framework would fit the *differential susceptibility* theory when you only consider positive environmental exposures (**Figure 1.2**). The direct test of either the *vantage sensitivity* model or the *differential susceptibility* model for liability to MDD would require high-quality measures of positive environmental exposures. Unfortunately, such data is limited in most population-based cohorts that have genetic data available, much more than measures of negative environmental exposures. Ideally, in the case of testing the *differential susceptibility* model, such measures must capture the full spectrum of positive and negative effects from environmental influences.

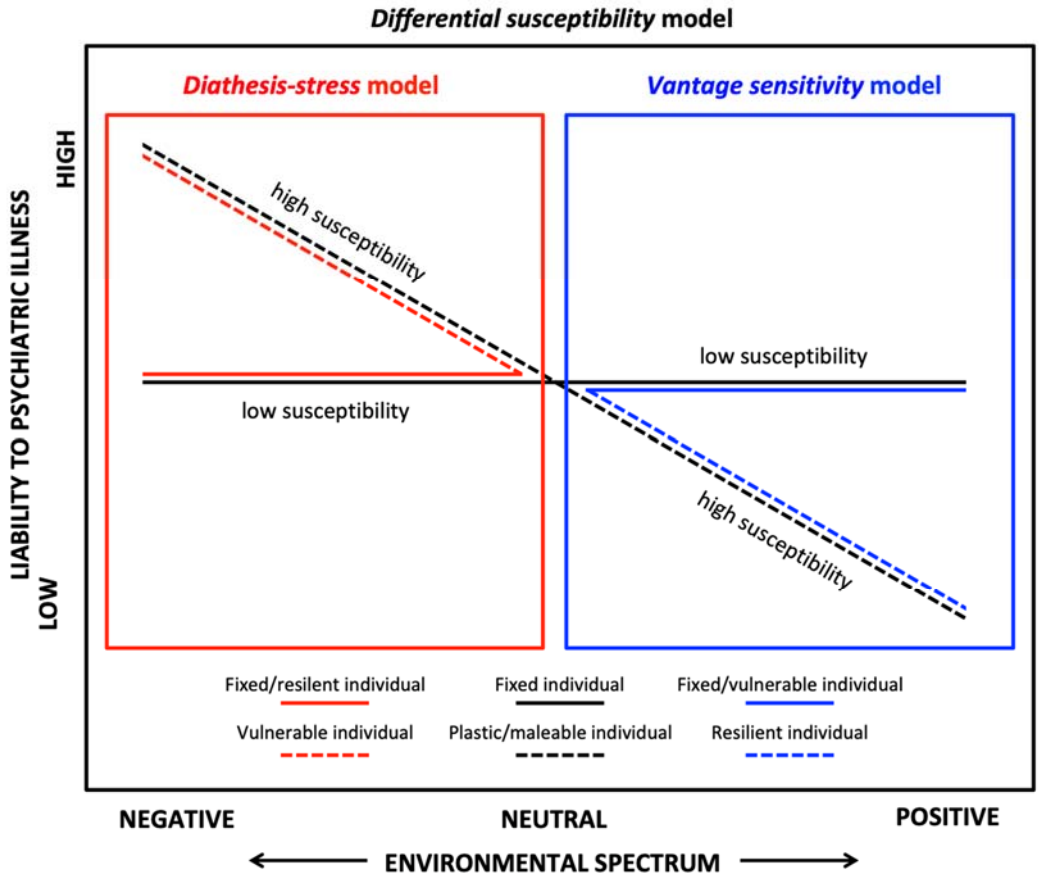


Figure 1.2 Representative illustration of: the diathesis-stress theory, the vantage sensitivity theory and the differential susceptibility theory. *In red, under the diathesis-stress model, individuals with high diathesis or vulnerability to illness are at higher risk of psychiatric illness when they are exposed to adverse and negative environments than resilient individuals with low diathesis. Conversely, under the vantage sensitivity model in blue, individuals with high susceptibility to the environment exposed to positive environments are at lower risk of psychiatric illness and more resilient than vulnerable individuals with low susceptibility. Under the differential susceptibility model in black, plastic individuals with high susceptibility to the environment will be more malleable and at higher risk to psychiatric illness in negative environments but at lower risk in positive environments than individuals with low susceptibility to environmental influences.*

1.6 Research on gene-by-environment interactions

If genes and the environment interact, gaining a good understanding of either genetic or environmental effects requires studying them in tandem. Research on GxE can help to detect new genetic effects associated with depression and to identify some of the underlying biological pathways involved[233,234]. In addition, the presence of a substantial GxE effect may, at least in part, help to explain some of the negative findings of main effects found in the literature. For example, as GWAS do not take into account environmental effects, those genetic variants that predispose to MDD under negative environments (i.e. that have a GxE effect) may be enriched in healthy individuals selected into control samples (some of which may did not develop symptoms because were exposed to neutral of positive environments). Thus, the power to detect association between these variants and MDD in a GWAS is dependent on the distribution of positive-negative environments between and within case-control samples. Therefore, the power to detect such variants in a GWAS would be maximised if we take into account environmental influences.

1.6.1 Gene-environment interaction (GxE)

Genetic variation can influence and modify the response to, and the effect of, environmental risk factors. Interpreted in another way, the effect of a genetic variant can be altered by an environmental exposure. This relationship is known as gene-by-environment interaction (GxE)[74,235]. In regard to this thesis, GxE can be defined as genetic differences in susceptibility to the effects of SLE, or stress causing differences to the genetic vulnerability.

There are two main types of GxE: the *fan-shaped* interaction and the *crossover* interaction[236]. Both types are illustrated in **Figure 1.3**. The *fan-shaped* interaction is the underlying GxE in depression hypothesized by the *diathesis-stress* theory. Under such interaction, the risk of depression depends on both the genotype and the environment with a genetic control of

mean liability to illness and sensitivity to the exposure. Note that this mode of interaction is a combination of the 1st and 2nd basic models from Kendler and Eaves. The *fan-shaped* interaction is the type of interaction assumed in most studies of stress and depression, and includes the known main effects of genetics and stress. Therefore, in accordance with a *fan-shaped* interaction, under stress, genetically vulnerable individuals are at higher risk of developing depressive symptoms than resilient individuals. However, in the absence of stress, such risk is not significantly different among individuals. The *fan-shaped* is the type of interaction underpinning the difference in estimates of heritability by environmental context detected in twin studies[237-239]. Conversely, the *crossover* interaction supports the *Differential Susceptibility* model and the concept of genetic plasticity rather than vulnerability. Therefore, in accordance with a *crossover* interaction, individuals that are more prone to develop symptoms under stressful and adverse environments will be as well more resilient under positive and favourable environments[230]. This type of interactions is consistent with evolutionary developmental theories that support that such variability in environmental response would be beneficial under selection pressure; and related to the theory of biological sensitivity to context[240]. It has been reported that children genetically prone to be highly reactive to stress display significantly higher rates of morbidity under stressful environments but significantly lower rates under supportive environments[241,242]. Unlike *fan-shaped* interactions, *crossover* interactions do not assume a main effect of genetics and environment, as they could, theoretically, have equal distributions between levels of exposure (see bottom right plot in **Figure 1.3**). However, this specific type of *crossover* interaction, if existent, must be rare[236]. *Crossover* interactions require less power to be detected than *fan-shaped* interactions, however, in practice they are difficult to distinguish from each other. As the crossover point of the interaction can vary along the exposure spectrum, both *fan-shaped* and *crossover* interactions can be present depending on the spectrum of the exposure under study.

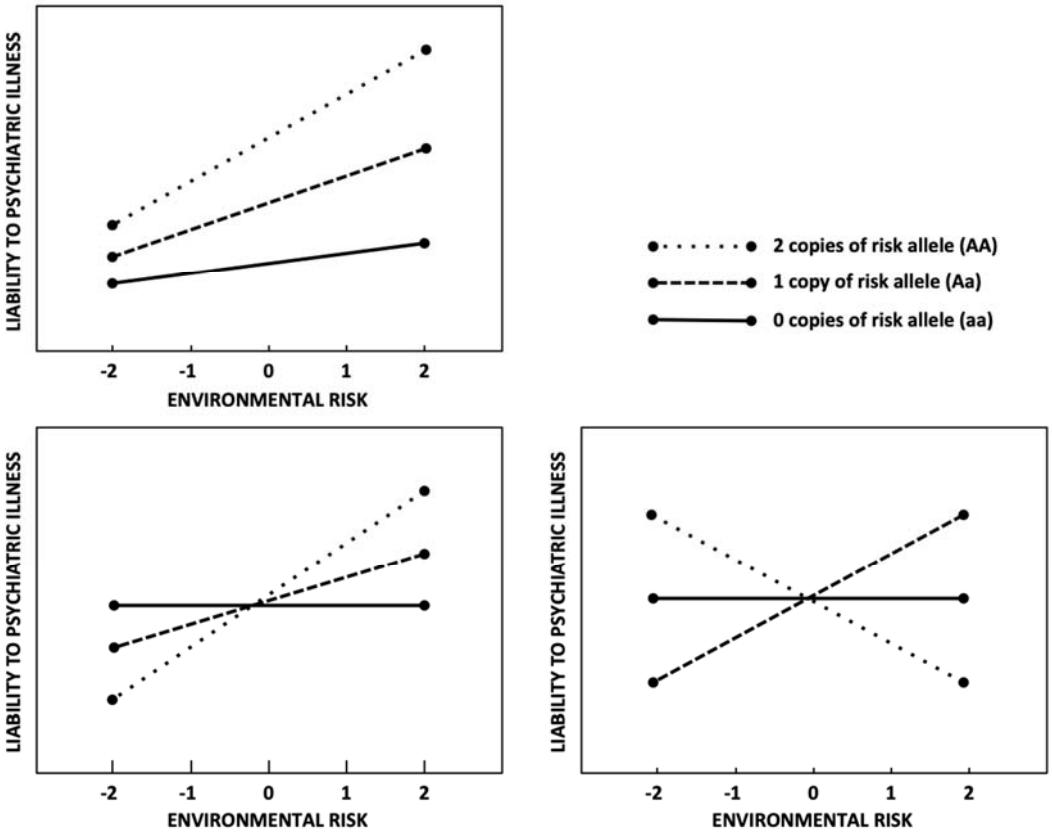


Figure 1.3 Types of gene-by-environment interactions. *Fan-shaped (top row) and crossover (bottom row) interactions. In a fan-shaped interaction, the environment itself has an effect on depression. See that the slope of the lines for each risk allele is continuous, indicating that the genetic risk on depression increases homogeneously by increment of exposure. Conversely, in a crossover interaction main effects are not assumed. See that the environmental effect would be detected only when the mean risk is different along the environmental spectrum (left) but not when distributions were equal (right). Note that if we expand the environmental spectrum (if we expand the lines) at some point we would see a crossover point and thus, such fan-shaped interaction would become a crossover interaction. Similar, if we only consider positive or negative influence from the environment in a crossover interaction with crossover point in neutral environment, we would detect a fan-shaped interaction. Plots adapted from Dick 2011*[236].

1.6.2 Gene-environment correlation

Not all forms of gene-environment interplay are considered GxE[235]. As shown previously by the 3rd model from Kendler & Eaves, exposure to specific environmental conditions can partly depend on an individual's genotype, so instead to be influenced by an independent and random environment, it may be driven by genetic control (e.g. through specific behaviours or personality traits such as neuroticism)[74,108]. This relationship is known as gene-environment correlation[243]. Gene-environment correlation can be categorizes as: *passive*, *active* and *evocative*. *Passive* gene-environment correlation occurs when an individual is exposed to environments established by their parents genetically influenced behaviour. *Active* gene-environment correlation occurs when an individual is exposed to environments selected by its own genetically influenced traits. For example, the gene-environment interplay seen in the 3rd model proposed by Kendler & Eaves of genetic control of exposure to environment is an example of active gene-environment correlation, but it is not a case of GxE. Finally, *evocative* gene-environment correlation occurs when an individual is exposed to environments or environmental responses from others that are evoked by his/her genetically influenced behaviour. As Kendler and Eaves suggested, genes may partially act by altering the environment to which individuals are exposed. This is also known as "nature via nurture", and is an example of how genes can influence the final outcome via environmental exposure. However, this does not mean that the effect of such environment is moderated by the genotype, nor the effect of the genotype itself moderated by the environmental exposure. We must remember that correlation does not imply causation and distinguish a pure GxE from gene-environment correlation. Below, I show two similar cases to further exemplify the difference between GxE and gene-environment correlation.

Case of GxE:

Imagine that living in poverty is a risk factor for depressive symptoms. Imagine that if carriers of allele "X" at a polymorphic locus live in poverty at

young ages, they are at a higher risk of suffering depressive symptoms in adulthood. However, if carriers of allele “X” do not live in poverty, they are resilient to developing depressive symptoms. Whilst, non-carriers of allele “X” at such locus tend to be resilient regardless of whether they lived in poverty at young age or not. This is a case of a GxE between allele “X” and living in poverty at young age on odds of suffering depressive symptoms in adulthood. This is a case where the effect of the environment is moderated by the genotype, or the environment moderates the effect of the genotype.

Case of gene-environment correlation:

However, now imagine that the odds of suffering depressive symptoms as a result of exposure to negative environments are the same for a group of people, but carriers of allele “Z” at a polymorphic locus behave in a way that contributes to be exposed to “negative” environments (e.g. poverty) at young ages at a higher rate than non-carriers of allele “Z”. Then, due to have been exposed to “negative” environments at a young age, carriers of allele “Z” have a higher rate of depressive symptoms in adulthood. This is a case of gene-environment correlation. Allele “Z” does not modulate the effects of living in poverty on odds of suffering depressive symptoms.

1.6.3 Statistical framework to test GxE

GxE studies, from primary GxE studies on candidate genes to the recently emerging genome-wide by environment interaction studies (GWEIS), test whether the joint effect of both genetic contributions and environmental risk factors is significantly different from the additive product of their individual effects. The statistical method to test for GxE effects differs from that used to test main additive effects in standard GWAS, which consist of a regression model (e.g. logistic or linear) that test whether the effect of a genotype (β_1) differs significantly from 0 (note that such test under a logistic model with absence of covariates would be roughly equivalent to an allele counting chi-square test in a case-control sample):

$$y_i = \beta_0 + \beta_1 G_i + \epsilon_i$$

The statistical method to test GxE effects in GxE studies consist of a similar regression model (e.g. logistic or linear) that additively fits the direct main effects of both the genotype and the environmental exposure plus the product of both terms (the interaction effect). The basic equation to study GxE is as follows:

$$y_i = \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i x E_i + \epsilon_i$$

In such an equation, we have y_i representing the phenotype under study (e.g. MDD status or depressive symptoms) as a function of the effects (β) from an individual's genotype or genetic variable (G_i; e.g. allele, genotype or genetic score such as PRS), the environmental variable (E_i; e.g. number of SLE reported in the last 6 months), the interaction between both factors ($G_i x E_i$) and an error term ϵ_i, which may include the effect of control variables (covariates) such as sex, age or others related to the sample under study. To assess whether there is a substantial GxE effect (β_3), the following hypothesis is tested:

$$H_0: \beta_3 = 0 \; vs. \; H_1: \beta_3 \neq 0$$

Thus, if $\beta_3 \neq 0$, it means that the interaction between "genetics" (G_i) and the environment (E_i) significantly contribute on the output. Therefore, if MDD status were the output, it would mean that there is a significant GxE effect underlying the aetiology of MDD.

An alternative test is to jointly assess the contribution of the genotype and the GxE together[233,244]. This is called a joint test and assesses the combined main additive G_i and $G_i x E_i$ effect on the output. It compares the model above to study GxE against a null model where both β_1 and β_3 equal 0. Therefore, to test the significance of a joint effect, the following hypothesis is tested:

$$H_0: \beta_1 + \beta_3 = 0 \; vs. \; H_1: \beta_1 + \beta_3 \neq 0$$

This can be conceptualized as the total genetic effect ($\beta_1 + \beta_3$).

1.6.4 GxE studies: transition from classical approaches and candidate genes to whole-genome studies

Early evidence for the presence of non-additive genetic effects on depressive disorders were found in studies based on structural models using twin, pedigree or adoption data, without requiring genetic data[245]. These studies used the relatedness between individuals to estimate a proxy of the genetic risk expected for each individual. Combining this latent measure of genetic influence with the environmental context derived from the data structure, these studies provided a range of ways to infer potential GxE in depression. They compared how genetic and environmental influences change across groups with different risk proxies, suggesting that the association between environmental stress and depression was partially mediated by familiar contributions including both genetics and shared environment[93,103,116,246-248].

1.6.4.1 Candidate gene-by-environment interaction studies

Initial implementations of genetic data to investigate GxE effects were candidate gene studies. In early 2002, Caspi *et al*. reported that maltreated children carrying a functional polymorphism associated with higher activity of the neurotransmitter-metabolizing enzyme monoamine oxidase A (MAOA gene) were less likely to develop antisocial disorders[249]. The hypothesis-driven strategy implemented was typical of the majority of following GxE studies on liability to depression. Probably the most influential and controversial GxE study on depression, performed by Caspi *et al.* in 2003, studied a serotonin transporter gene (*SLC6A4*; also known as *5-HTT* gene) as candidate gene. They reported that the likelihood of have suffered a MDE, depressive symptoms or attempted suicide in the past year was higher in carriers of one or two copies of the short allele from a functional polymorphism on a promoter region linked to the *5-HTT* gene (5-HTTLPR) than for homozygous carriers for the long allele[250]. This was a result of a dose-response interaction between 5-HTTLPR and the number of SLE reported over the preceding 5 years, and supported the hypothesis that

genetic variation can modulate an individual's sensitivity to environmental stress. Similar findings were reported when childhood maltreatment was considered as the environmental exposure. However, the robustness of these findings is still debated[92,251-254]. In fact, there are a large number of candidate gene GxE studies (more than fifty just attempting replication of Caspi's findings) reporting results of which is not clear how robust and replicable they are. Many other candidate gene studies have been performed with other polymorphisms from a restricted set of hypothesis-driven candidate genes, based on previous empirical evidence about the link between their biological functions and some environmental influences, e.g. *SLC6A4*, *FKBP5*, *CRHR1*, *COMT*, *CNR1, BDNF* or *MAPK14*[255-260]. These have mostly focused on the monoaminergic systems (e.g. serotoninergic, dopaminergic) or stress-response pathways linked to the HPA axis. Nevertheless, robust findings have been absent and the results are in doubt due to their lack of power[254]. Two meta-analyses concluded, a few years after Caspi's findings, that there was no evidence of a significant GxE effect between the 5-HTTLPR polymorphism and SLE in depression in any sex or at population level[251,252]. However, Caspi *et al.* came back with a later review emphasizing that evidence for GxE effects come from a wide range of studies including observational studies in humans and non-human primates, studies of *5-HTT* mutations in rodents and from experimental neuroscience studies[261]. The presence of such GxE effect was later support by a larger meta-analysis[253]. Nevertheless, a recent collaborative meta-analysis based on 31 studies including 38,802 individuals from European ancestry concluded that there was no evidence of a significant GxE, but that if such interaction exist, its effect must be modest and only detectable in specific scenarios[262]. Therefore, more than 15 years later, the debate about whether GxE between the serotonin transporter genotype and environmental stress increase liability to depression or not still remains open, highlighting how complex is the study of GxE effects.

1.6.4.2 Genome-wide GxE studies

Genome-wide GxE studies may follow different study designs (e.g. case-only, two-phase case–control or counter-matched designs)[263,264]. In 2014, the first genome-wide GxE study in psychiatric disorders was conducted in schizophrenia[265]. It tested a final set of 29,082 SNPs displaying nominal association with maternal load of cytomegalovirus infection with maternal load of cytomegalovirus infection as the "environmental" exposure. Although it did not achieve the accepted genome-wide significance threshold from standard genome-wide studies ($p = 5 \times 10^{-8}$) and it was not replicated, they reported a significant GxE ($p = 7.3 \times 10^{-7}$; Bonferroni significance $p = 1.72 \times 10^{-6}$) in *CTNNA3*, a gene previously unknown in schizophrenia[265]. If such association were replicated[266], this study would provide evidence for the importance of including environmental risk factors in genetic studies in order to identify genetic risk variants. In regards to liability to MDD, only a very few genome-wide GxE studies have been performed to date[113,267-270]. Recently, few studies have applied similar approaches to pre-select candidate polymorphism before testing for GxE effects on MDD. In 2017, Van der Auwera *et al.* selected 27 candidate genes and 268 SNPs previously associated with either MDD or GxE effects in MDD and tested for GxE effects using case-control and case-only designs in a PGC sample of 3,944 individuals of European ancestry[270]. No genome-wide significant finding was reported. Finally, Coleman *et al.* recently used over 73,000 participants from UK Biobank and a set of 1,652 imputed SNPs nominally associated ($p < 1 \times 10^{-4}$) in GWAS of depression and/or trauma exposure in order to investigate GxE effects in MDD status using a binary variable of overall self-reported trauma in childhood and adulthood[269]. They reported evidence of GxE with trauma exposure in 78 SNPs ($p < 3 \times 10^{-5}$).

Recently, a new strategy has emerged to investigate GxE at the genome-wide scale, the performance of genome-wide by environment interaction studies (GWEIS; sometimes called gene-environment-wide interaction studies)[263,271]. Paradoxically, nowadays this kind of strategy is more limited by lack of high-quality assessment of environmental factors, and the

existence of such data on “environmental” measures in general, than by availability and quality of genomic data. GWEIS refer to genome-wide GxE studies with epidemiological cohort and case–control designs that independently test genotyped SNPs along the genome without requiring any prior knowledge about the underlying biology of MDD. GWEIS can also test associations with the combined additive and GxE effects (i.e. the joint effect; further details in section 1.6.3)[233,244]. In 2016, Dunn *et al*. conducted the first and largest GWEIS of depressive symptoms to date in two samples of women from minority populations: 7,179 African American and 3,138 Hispanic/Latinas women[113]. They reported a significant GxE in the African American sample 14kb from *CEP350* (p = 4.10 x 10^{-10}), but this interaction was not replicated in independent samples of either African American or Hispanic/Latinas women. That same year, two other GWEIS on depressive symptoms in the Japanese population were published. Otowa *et al.* reported a significant GxE 8kb upstream of *RGS10* (p = 4.5 x 10^{-8}) in a cohort of 320 Japanese[267]. However, the veracity of such finding is very low due to its lack of power, which themselves reported as 2% in order to detect significant interactions accounting for at least 1% of the variance in such small sample size. Ikeda *et al.* conducted the other GWEIS in a sample of 1,112 Japanese individuals (91% women) using a robust joint test[268]. This robust joint test combines the effect of the GxE with the main SNP additive effect into a joint effect that may help to identify locus associated with MDD[272]. Using a measure of SLE, they reported a single significant association with such joint effect (p = 8.2 x 10^{-9}) between depressive state and a variant downstream of *BMP2*, a gene with no direct relationship with mood disorder susceptibility but with neurotrophic effects and enriched in neurons. Nevertheless, they did not test the interaction effect alone. Noteworthy, the robust joint test implemented on this GWEIS adjusts by a systematic and spurious genome-wide inflation that occurs on most GxE studies due to heteroscedasticity in the data. Heteroscedasticity is a statistical phenomenon that arises when the variance of the trait under study (e.g. depressive symptoms) is not homogeneous across each level of the environmental exposure (e.g. number of reported

SLE). Therefore, this phenomenon must be taken into account in GWEIS. Ikeda *et al.* conducted such GWEIS applying an R plugin for PLINK[273] developed by Almli *et al.*[272].

1.6.4.3 Polygenic risk scores to test GxE effects

A rising alternative to investigate GxE effects locus by locus is to use PRS weighted by genetic factors associated with the phenotype of interest and test their interaction with an environmental variable (i.e. PRSxE)[274]. PRS explain more phenotypic variance and have much greater explanatory power than individual SNPs. In addition, by aggregating all weightings from genome-wide data into a single score there is only a single genetic independent variable to be tested and thus, it reduces the multiple testing. This new approach has been applied to investigate GxE effects on MDD with childhood trauma[83,86,275] and adult SLE[86,169,276]. However, initial results contradict. In 2014, Peyrot *et al*. conducted a study in European individuals involving 1,645 MDD cases and 340 screened controls[83] using PRS weighted by main additive effects of MDD derived from a PGC genome-wide meta-analysis[147]. They reported that individuals with high genetic vulnerability were at higher risk of developing MDD when they had been exposed to childhood trauma, and that the risk increased with increasing severity of the events. In a similar approach, and using the same GWAS summary data, in a US population-based sample of 8,761 older adults, Musliner *et al*. reported no significant interaction between PRS and recent SLE on current depressive symptoms[276]. They therefore concluded that there was no evidence of SLE modulating polygenic risk. Later, a replication study from Mullins *et al*. assessed the interaction of the same PRS in a UK sample involving 1,605 individuals with MDD and 1,064 controls[86]. In this case, they assessed PRS interaction with both childhood trauma and adult SLE reporting similar findings to the previous studies of Peyrot *et al*. and Musliner *et al*. but with controversial interpretations for the modulating effects of childhood trauma (i.e. opposing interaction effect). As in Musliner *et al.*, they found no significant interaction between PRS and adult SLE, but significant interaction between PRS and childhood trauma, suggesting that only very severe events

such as childhood trauma modulated the risk effect, or that only the effect of very severe events such as childhood trauma is modulated by genetic variation. Furthermore, contrary to the conclusions of Peyrot *et al.*, individuals who suffer moderate to severe childhood trauma were found to be at higher risk of MDD when they had lower genetic risk of MDD. Conversely to both studies, a recent meta-analysis conducted by Peyrot *et al*., which included both studies investigating childhood trauma as exposure along with several other cohorts contributing to the PGC and involved 3,024 MDD cases and 2,741 controls, reported no significant interaction between PRS (weighted by a substantially larger discovery sample of ~110,000 individuals) and childhood trauma. Therefore, concluding that previous evidence from Peyrot *et al*. and Mullins *et al*. were no longer significant when using a PRS weighted by a larger discovery sample and thus, they should be interpreted as chance findings. Although further research is required, the most recent evidence suggests that genetics does not modulate the negative effect of childhood trauma on odds of suffering depression in adulthood.

Conversely, Colodro-Conde *et al.* recently tested the interaction between PRS and adult SLE (i.e. two kinds of SLE, so-called personal and network SLE), as well as the interaction with social support (i.e. a measure of positive environment), in a sample of 5,221 individuals from 3,083 Australian families following a *diathesis-stress* framework[169]. Unlike the previous studies described, they used a genomic relationship matrix and a linear mix model implemented in GCTA to estimate the variance in depression explained by the PRSxE while taking into account the independent effects of PRS and environment scores. They used PRS weighted by the most recent genome-wide meta-analysis conducted by PGC at that time (N=159,601) and more extensive and informative measures of SLE. In contrast to the previous Musliner *et al*. and Mullins *et al*. studies, they reported a significant interaction between PRS and SLE on liability to depression, using personal SLE and mainly driven by effects in women. Thus, concluding that there is an extra risk on liability to depression in individuals that combine high genetic

risk of MDD and high number of reported SLE. They found no significant interaction with network SLE or social support.

1.7 Thesis summary and aims

The aim of this thesis is to advance our understanding of the genetic responses to environmental stress underlying MDD. Herein, I present findings from a range of new genome-wide studies and polygenic approaches, using data from Caucasian populations of United Kingdom ancestry, in order to investigate the complex interplay between genetic factors and environmental stress in the aetiology of depression.

In **chapter 2** and **chapter 3** I explore the genetic basis of a proxy for sensitivity to stress without direct measures of environmental stress but exploiting the relationship between stress, neuroticism and depression. In **chapter 2**, such proxy for sensitivity to stress is operationalized through a genome-wide interaction study, and **chapter 3** explores the enrichment of genetic factors contributing to risk of MDD and stress sensitivity within the HPA axis and glucocorticoid pathways. In **chapter 4**, **chapter 5** and **chapter 6** I incorporate self-reported quantitative measures of recent SLE to investigate the genetic response to SLE in depression. **Chapter 4** empirically tests the *diathesis-stress* model for depression, using PRS weighted by the additive effects of MDD derived from the PGC MDD GWAS and SLE measures, in order to validate previous findings supporting the *diathesis-stress* theory. **Chapter 5** presents the results from GWEIS conducted in two cohorts that seek to identify common variants with GxE effect on depression in response to SLE. Finally, **chapter 6** incorporates the weightings for genetic stress sensitivity and stress response derived from **chapters 2** and **chapter 5** into a new assessment of the *diathesis-stress* model for depression following the *diathesis-stress* framework implemented in **chapter 4**. Finally, in **chapter 7** I discuss my findings, some limitations and pitfalls of this research area, and the future perspective for GxE studies underlying MDD.

Chapters 2, **chapter 4** and **chapter 5** have been accepted for publication to peer-reviewed journals. As of January 2019, **chapter 2** has been published in *PLOS ONE* (attached into the **Appendix A.4**), and both **chapter 4** and **chapter 5** have been published in *Translational Psychiatry* (attached into the **Appendix C.3** and the **Appendix D.7**, respectively). **Chapter 6** is going to be submitted to *PLOS Genetics*. All the studies presented in this thesis are my own work based on my research conducted during the 4 years of my PhD program in Computational Biology.

Chapter 2 Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder

Well-powered GxE studies require having a large sample with both genetic and environmental data available from the same individuals. However, data on key environmental exposures such as SLE or childhood maltreatment is missing or limited by a lack of high quality in most population-based cohorts with genetic data available. The quality of measures on environmental stress tends to decline in relation to sample size and even fewer prospective studies exist with this data available. However, several population-based cohorts with genetic data have measures on neuroticism levels. In this chapter, I exploit the relationship between environmental stress, neuroticism and MDD to derive a proxy for stress sensitivity without requiring direct measures of SLE.

This chapter is presented as it has been published in PLOS ONE, which explains the use of "we" within the chapter. I confirm that the work presented (i.e. investigation, conceptualization, data curation, formal analysis, plot generation and writing) is my own work under guidance from my supervisor Dr. Pippa Thomson. She provided the original idea and the initial draft of the introduction. Co-authors contributed with the collection, access and preliminary curation of some raw data and/or providing critical revisions. I performed all the analyses myself. The published article and **Supplementary Material** can be found in **Appendix A**.

Publication:

Arnau-Soler, A. *et al.* Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS One* **13**, e0209160 (2018).

2.1 Abstract

Individual response to stress is correlated with neuroticism and is an important predictor of both neuroticism and the onset of major depressive disorder (MDD). Identification of the genetics underpinning individual differences in response to negative events (stress-sensitivity) may improve our understanding of the molecular pathways involved, and its association with stress-related illnesses. We sought to generate a proxy for stress-sensitivity through modelling the interaction between SNP allele and MDD status on neuroticism score in order to identify genetic variants that contribute to the higher neuroticism seen in individuals with a lifetime diagnosis of depression compared to unaffected individuals. Meta-analysis of genome-wide interaction studies (GWIS) in UK Biobank (N = 23,092) and Generation Scotland: Scottish Family Health Study (N = 7,155) identified no genome-wide significance SNP interactions. However, gene-based tests identified a genome-wide significant gene, *ZNF366*, a negative regulator of glucocorticoid receptor function implicated in alcohol dependence (p = 1.48 x 10^{-7}; Bonferroni-corrected significance threshold p < 2.79 x 10^{-6}). Using summary statistics from the stress-sensitivity term of the GWIS, SNP heritability for stress-sensitivity was estimated at 5.0%. In models fitting polygenic risk scores of both MDD and neuroticism derived from independent GWAS, we show that polygenic risk scores derived from the UK Biobank stress-sensitivity GWIS significantly improved the prediction of MDD in Generation Scotland. This study may improve interpretation of larger genome-wide association studies of MDD and other stress-related illnesses, and the understanding of the etiological mechanisms underpinning stress-sensitivity.

2.2 Introduction

Stressful life events are known to increase liability to mental illness and disease-related traits[277] including neuroticism[218-220], major depressive disorder (MDD)[79,84,85], autoimmune diseases[278] and some cancers[279,280]. A greater understanding of the causal mechanism by which negative events affect disease risk or outcome may be beneficial in identifying individuals for targeted support. However, it has been proposed that sensitivity to stress may be an important predictor of response to stress[188,281]. In particular, the effect on an individual may result more from the perceived stress than the event itself, and may be dependent on individual differences in stress-sensitivity[282-287]. Studies of *5-HTT* and twin studies suggest that stress-sensitivity may, at least in part, be heritable[261,288-290]. Despite a complex interaction between MDD, neuroticism and stress, multivariate structural equation models have confirmed a genetic effect on perceived stress, overlapping that on MDD or neuroticism, but with a specific genetic component[288]. The inter-relatedness of these traits may offer an approach to identify the genetic variation that affects an individual's stress-sensitivity, and improve genetic prediction of an individual's liability to negative outcomes. By modelling the interaction between SNP allele and MDD status on neuroticism score through genome-wide interaction studies (GWIS), we sought to investigate the genetics of stress-sensitivity.

The personality trait neuroticism is moderately heritable (30–50% estimates from twin studies)[189,203,204,291], is higher in individuals with depression compared to controls[213,214] and is known to have shared genetic aetiology with depression[292-295]. Neuroticism is strongly correlated with measures of sensitivity to punishment but not reward[296], positively correlated with perceived personal relevance of a stressor[297,298] and has been used previously as a proxy measure of stress-sensitivity[221]. Neuroticism is thought to mediate or interact with the effects of adverse life events on risk of depression[79,222]. It has a substantial stable component[299], however, there is

evidence for change, as well as stability, across the life span[217-220]. Individual differences in neuroticism are enduringly influenced by both genetic and environmental factors[300]. Whereas the stable component of neuroticism is strongly determined by genetics, change in neuroticism score is attributed to the effects of unshared environment[217]. Persistent change in neuroticism score has been shown in response to life events[218-220]. Negative life events lead to small persistent increases in neuroticism over time[220]. However, recent stressful life events (β = 0.14 95%CI 0.13 - 0.15, $p < 0.001$) have a stronger effect than distant stressful life events suggesting a reduction of effect over time[220]. Long-lasting increases in neuroticism associated with distant negative life events are mediated by depression[218].

Major depressive disorder (MDD) is a complex disorder influenced by both genetic contributions and environmental risk factors, with heritability estimates from twin and family studies of between 31-42%[119,150]. Confirmed environmental risk factors for MDD include maternal infections, childhood maltreatment and negative life events[79,81,83-85]. However, few genetic studies have such information and even fewer prospective studies exist. Incorporation of stressful life events has been shown to improve the ability to predict MDD[86,169] and, although stress is an environmental risk factor, it may have an independent genetic contribution to risk of depression[97,103,114,115,169].

These studies suggest that a genetic variable derived from the difference in neuroticism levels seen in individuals with MDD compared to controls may allow us to identify genetic loci important for stress-sensitivity. We sought to identify the genetic underpinnings of individual's sensitivity to stress response (stress-sensitivity) by identifying variants that contribute to the higher neuroticism levels seen in individuals with a lifetime diagnosis of MDD. Further, polygenic risk scores (PRS) derived from this stress-sensitivity variable may improve prediction of MDD over that based on MDD or neuroticism PRS alone.

Using unrelated individuals from two large population-based samples, UK Biobank (UKB; N = 23,092) and Generation Scotland: Scottish Family Health

Study (GS:SFHS; N = 7,155), we sought to identify genes involved in stress-sensitivity by performing GWIS for the interaction between MDD status and SNP allele on neuroticism score. We identified a gene significantly associated with stress-sensitivity and show that a PRS derived from the interaction term of the GWIS, significantly predicts liability to depression independently of the PRS for MDD and/or neuroticism.

2.3 Materials and methods

2.3.1 UK Biobank (UKB) Participants

UKB is a major national health resource that aims to improve the prevention, diagnosis and treatment of a wide range of illnesses. It recruited more than 500,000 participants aged from middle to older age who visited 22 assessment centres across the UK between 2006 and 2010. Data were collected on background and lifestyle, cognitive and physical assessments, sociodemographic factors and medical history. The scientific rationale, study design, ethical approval, survey methods, and limitations are reported elsewhere[301,302]. UKB received ethical approval from the NHS National Research Ethics Service North West (Research Ethics Committee Reference Number: 11/NW/0382). All participants provided informed consent. The present study was conducted on genome-wide genotyping data available from the initial release of UKB data (released 2015). Details of sample processing specific to UKB project are available at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583 and the Axiom array at http://media.affymetrix.com/support/downloads/manuals/axiom_2_assay_auto_workflow_user_guide.pdf. UKB genotyping and the stringent QC protocol applied to UKB data before it was released can be found at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580. SNPs genotyped on GS:SFHS were extracted from the imputed UKB genotype data[303] (imputed by UKB using a merged panel of the UK10K haplotype reference panel and the 1000 Genomes Phase 3 reference panel) with quality > 0.9 was hard-called using PLINK v1.9[273]. Individuals were removed based on UKB genomic analysis exclusion (UKB Data Dictionary item #22010), non-white British ancestry (#22006: genetic ethnic grouping; from those individuals who self-identified as British, principal component analysis was used to remove outliers), high genotype missingness (#22005), genetic relatedness (#22012; no pair of individuals have a KING-estimated kinship coefficient > 0.0442),

QC failure in UK BiLEVE study (#22050 and #22051: UK BiLEVE Affymetrix and UK BiLEVE genotype quality controls for samples) and gender mismatch (#22001: genetic sex). Further, from the initial release of UKB data and using PLINK pi-hat < 0.05, individuals who were also participants of GS:SFHS and their relatives were excluded to remove any overlap of individuals between discovery and target samples. A dataset of 109,283 individuals with 557,813 SNPs remained for further analysis, aged 40-79 (57,328 female, 51,954 male; mean age = 57.1 years, s.d. = 7.99), of which 109,282 had data available for neuroticism score and 23,092 had data available on MDD status (n_{cases} = 7,834, $n_{controls}$ = 15,258, n_{female} = 11,510, n_{male} = 11,582; mean age = 57.7 years, s.d. = 8.04). Thus, the final dataset comprised 23,092 unrelated individuals.

2.3.2 Generation Scotland Scottish Family Health Study (GS:SFHS) Participants

GS:SFHS is a family-based genetic epidemiology study which includes 23,960 participants from ~ 7,000 Scottish family groups collected by a cross-disciplinary collaboration of Scottish medical schools and the National Health Service (NHS) from February 2006 to March 2011. Participants were interviewed and clinically assessed for a wide range of health-related traits (including high-fidelity phenotyping for Major Depressive Disorder and related endophenotypes), environmental covariates and linked to routine health records[304,305]. All components of GS:SFHS obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the NHS (Research Ethics Committee Reference Number: 05/S1401/89) and participants provided written consent. The protocol for recruitment is described in detail in previous publications[124,306]. GS:SFHS genotyping and quality control is detailed elsewhere[307]. Briefly, individuals with more than 2% missing genotypes and sex discrepancies were removed, as well as population outliers. SNPs with genotype missingness > 2%, minor allele frequency < 1% and a Hardy-Weinberg Equilibrium test $p < 1 \times 10^{-6}$ were exclude. Finally, individuals were removed based on relatedness (pi-hat < 0.05), maximizing retention of case individuals, using PLINK v1.9[273].

Genome-wide SNP data for further analysis comprised 7,233 unrelated individuals genotyped for 560,698 SNPs (n_{female} = 3,476, n_{male} = 3,757), aged 18-92 (mean age = 50.4 years, s.d. = 12.06) of which: 7,190 had clinical data on MDD; 7,196 individuals had data on neuroticism; and 7,155 had data on both neuroticism and MDD.

2.3.3 Phenotype assessment

2.3.3.1 Neuroticism score (EPQN)

Participants in both UKB and GS:SFHS cohorts were assessed for neuroticism using 12 questions from the Eysenck Personality Questionnaire-Revised Short Form's Neuroticism Scale (EPQN)[198,308-310]. Neuroticism can be scored by adding up the number of "Yes" responses on EPQN. This short scale has a reliability of more than 0.8[201]. EPQN distributions were found to be sufficiently "normal" after assessment for skewness and kurtosis to be analysed using linear regression (both coefficients were between -1 and 1).

2.3.3.2 MDD diagnoses

In UKB, the MDD phenotype was derived following the definitions from Smith et al.[310] Current and previous depressive symptoms were assessed by items relating to the lifetime experience of minor and major depression[308], items from the Patient Health Questionnaire[311] and items on help-seeking for mental health[310]. Using a touchscreen questionnaire, participants were defined as probable cases if they i) answered "Yes" to the question "Ever depressed for a whole week" (UKB field: 4598), plus at least 2 weeks duration (UKB field: 4609), or ii) did report having seen a GP or psychiatrist for nerves, anxiety, tension or depression (UKB fields: 2090 and 2010) and reported symptoms (UKB field: 4631) with at least 2 weeks duration (UKB field: 5375). In our unrelated sample, 7,834 participants were diagnosed with MDD (with single, moderate or recurrent episodes) and 15,258 were controls (N = 23,092).

In GS:SFHS, participants took in-person clinical visits where they were screened for a history of psychiatric and emotional disorders (i.e., psychiatric, mood state/psychological distress, personality and cognitive assessment) by

trained researchers using the Structured Clinical Interview for DSM-IV Non-Patient Version (SCID)[312], which is internationally validated to identify episodes of depression. Those participants that were positive in the initial screening continue through clinical interview and were administered the mood sections of the SCID. The SCID elicited the presence or absence of a lifetime history of MDD, age of onset and number of episodes. Participants fulfilling the criteria for at least one major depressive episode within the last month were defined as current MDD cases. Participants who were screened positive for Bipolar I Disorder were excluded. Those participants who were negative during the initial screening or did not fulfilled criteria for MDD were assigned as controls. Further details regarding the diagnostic assessment are reported elsewhere[124,305]. All interviewers were trained for the administration of the SCID. Inter-rater reliability for the presence or absence of a lifetime diagnosis of major depressive disorder was good (Kappa = 0.86, $p < 0.001$, 95%CI 0.7 to 1.0). In our unrelated GWIS sample (N = 7,155), 2,010 had a lifetime diagnosis of MDD and 5,145 were controls.

2.3.4 Statistical Methods

2.3.4.1 GWIS and derivation of a genetic stress-sensitivity effect

The effect size of an stress-sensitivity effect (β_{SS}) was derived by performing a GWIS for the effect of the MDD status and SNP allele on EPQN (dependent variable) in both UKB and GS:SFHS cohorts using PLINK 1.90 (PLINK-command --gxe; fitting MDD diagnosis as a binary "group" effect)[273]. PLINK-command --gxe estimates the difference in allelic association with a quantitative trait (EPQN) between two groups (MDD cases vs. controls) producing effect estimates on each group and a test of significance for the interaction between SNP allele and MDD status. The interaction *p* value reflects the difference between the regression coefficient of the allelic effect in a linear model for EPQN in MDD cases (β_A) and the same regression coefficient in a linear model for EPQN in controls (β_B). The stress-sensitivity interaction effect was defined as the difference in allele effect between MDD cases and control groups.

Considering one SNP, the effect it confers to EPQN can be modelled by MDD status (control = 0, MDD case = 1) as follows:

$$\begin{cases} MDD = 0;\ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_A SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

This is equivalent to modelling the effect on MDD cases as follows:

$$\begin{cases} MDD = 0;\ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_B SNP + (\beta_A - \beta_B) SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

Or, it can be modelled as a whole as:

$$EPQN = \beta_0 + \beta_2 MDD + \beta_B SNP + (\beta_A - \beta_B) SNP * MDD + \beta_{0c} COV + \beta_{2C} COV * MDD + \varepsilon$$

Where *COV* stands for covariates, β_2 stands for $\beta_1 - \beta_0$, and β_{2c} stands for $\beta_{1c} - \beta_{0c}$.

Thus, the interaction effect (β_{SS}) can be estimated as the difference in allelic effect on EPQN between MDD cases (β_A) and controls (β_B) as follows,

$$\hat{\beta}_{SS} = \hat{\beta}_A - \hat{\beta}_B$$

$\hat{\beta}_{SS}$ is therefore defined as the effect size reflecting the genetic stress-sensitivity effect on MDD cases compared to controls (**Appendix A**: Supplementary Figure 1).

2.3.4.2 Stress-sensitivity GWIS, main additive effect GWASs, meta-analysis and gene-set analysis.

For GWIS and subsequent analyses, sample specific covariates were applied as follows: *UKB.* All phenotypes were adjusted for centre, array and batch as random effects prior to analyses. Analyses were adjusted for age, sex and 15 informative principal components (PCs; UKB Data Dictionary items #22009.01 to #22009.15) as fixed effects to take account of possible population stratification. *GS:SFHS.* All the analyses were adjusted for age, sex and 20 PCs.

GWAS for MDD and neuroticism, using logistic and linear models of additive allelic effects respectively, were conducted on the same sample sets for comparison and generation of matched PRS using PRSice-2[313].

Results from the GWIS of UKB and GS:SFHS were combined in a sample size weighted meta-analysis performed using METAL[314]. While the use of standard error weighting is more common, the different diagnostic scheme and MDD prevalence between the two cohorts (GS:SFHS; 12.2%, UKB: 25.8%)[124,310] may indicate systematic differences in the measurement of MDD. Generalized gene-based analysis of the meta-analysis was performed using MAGMA[315] implemented through FUMA[316] (http://fuma.ctglab.nl). Briefly, SNP summary statistics were mapped to 17,931 protein-coding genes. Individual SNP *p* values from a gene were combined into a gene test-statistic using a SNP-wise model and a known approximation of the sampling distribution used to obtain a gene-based *p* value. Genome-wide significance was defined at $p = 0.05/17{,}931 = 2.79 \times 10^{-6}$.

2.3.4.3 LD Score regression

The summary statistics from the meta-analysis were used to examine the genetic overlap between the polygenic architecture of stress-sensitivity, MDD and neuroticism. LD score regression was used to derive the genetic correlations (r_G) between these traits[173,174] using meta-analysed GWAS and GWIS summary statistics. SNP-based heritability was also estimated using LD score regression, using the summary statistics from single-SNP analyses.

2.3.4.4 Pathway, functional and gene expression analyses

Lead SNPs, independently associated with the phenotype, were identified using PLINK 1.90 by clumping (*p* threshold $< 2 \times 10^{-5}$; LD r2 > 0.1; physical kb threshold = 500kb; 1000 Genomes Project Phase 1 CEU, GBR, TSI genotype data), and analysed using DEPICT[317]. Further detail is given in **Appendix A.1**.

Genes associated with lead SNPs were investigated for evidence of: phenotypic association in the NCBI dbGaP database of genotypes and

phenotypes[318] (https://www.ncbi.nlm.nih.gov/gap/phegeni), regulatory DNA elements in normal cell lines and association with expression quantitative trait loci (eQTLs) using the RegulomeDB database[319] (http://www.regulomedb.org) and the Genotype-Tissue Expression (GTEx) Portal[320] (http://www.gtexportal.org).

2.3.4.5 Polygenic profiling

PRS were produced using PRSice-2[313], permuted 10,000 times and standardized to a mean of 0 and a standard deviation of 1. Using GWIS summary statistics, we created PRS for stress-sensitivity (PRS_{SS}) by weighting the sum of the reference alleles in an individual by the stress-sensitivity effect (β_{SS}). Additional PRS were generated weighting by MDD main additive effects (PRS_D) and neuroticism main additive effects (PRS_N) using GWAS summary statistics from GS:SFHS or UKB. In addition, PRS_D and PRS_N were also generated using summary statistics from the most recent Psychiatric Genetic Consortium (PGC) MDD meta-analysis[150] (excluding GS:SFHS, and UKB individuals when required; N = 155,866 & 138,884) and the Genetics of Personality Consortium (GPC) neuroticism meta-analysis[204,321] (N = 63,661). Generalized linear models were implemented in R 3.1.3[322]. The direct effect of PRS_{SS} (model 1), PRS_D (model 2) and PRS_N (model 3) on MDD risk were assessed in independent logistic regression models on GS:SFHS (target cohort) using GWAS and GWIS statistics from UKB (the largest cohort) as the discovery sample to weight PRS. Multiple regression models fitting both PRS_D and PRS_N (model 4) and fitting each of them separately with PRS_{SS} (models 5 and 6) were also calculated. Finally, full additive multiple regression models fitting PRS weighted by all three effects (full model) was assessed using both PRS_{SS}, PRS_D and PRS_N at their best-fit in independent models. Further, results were also assessed using PRS_D and PRS_N weighted by PGC2 MDD[150] and GPC neuroticism[321] summary statistics. Further detail is given in **Appendix A.2**. All models were adjusted by sex, age and 20 PCs. A null model was estimated from the direct effects of all covariates on MDD. 10,000 permutations were used to assess significance of each PRS. The predictive improvement of

combining the effects of multiple PRS over a single PRS alone was tested for significance using the likelihood-ratio test.

Cross-validation was performed using UKB as target sample and GS:SFHS as discovery sample. Additional analyses using PRS_D and PRS_N weighted by PGC2 MDD[150] and GPC neuroticism[321] summary statistics were also tested. MDD status on UKB was adjusted by centre, array and genotyping batch as random effects and scaled (between 0 and 1) prior to analysis, giving a quasi-binomial distribution of MDD status on UKB. Models implemented on UKB (quasi-binomial regression) were adjusted by sex, age and 15 PCs. Nagelkerke's R^2 coefficients were estimated to quantify the proportion of MDD liability explained at the observed scale by each model and converted into R^2 coefficients at the liability scale (prevalence: 12.2% in GS:SFHS[124] and 25.8% in UKB[310]) using Hong Lee's transformation[172] available from GEAR: GEnetic Analysis Repository[323].

2.3.4.6 Using stress-sensitivity to stratify depression

GS:SFHS MDD cases (n_{cases} = 2,016; n_{female} = 1,345, n_{male} = 671) have data available on MDD course (single or recurrent), age of onset (n = 1,964) and episode count (n = 2,016), as well as on neuroticism (n = 2,010). In addition, a subset were evaluated by Mood Disorder Questionnaire[324] (MDQ; n = 1,022) and Schizotypal Personality Questionnaire[325] (SPQ; n = 1,093). The reduced sample number of MDQ and SPQ reflects the later addition of these questionnaires to the study and does not reflect a particular subgroup of GS:SFHS.

Difference in PRS_{SS} and PRS_D between MDD cases and controls on GS:SFHS were tested using a Student's two sample t-test (two tailed). Cases of MDD on GS:SFHS with data available on each trait analyzed were stratified by quintiles based on PRS_{SS} and PRS_D (5 x 5 groups). Post hoc, the effects on each trait of quintiles based on PRS_{SS} and its interaction effect with quintiles based on PRS_D were assessed using linear regression models adjusting by sex and age in an attempt to identify a characteristic subtype of

MDD patients with differential stress-sensitivity levels. The same analysis was reproduced using PRSs as continuous variables.

2.4 Results

We confirmed the elevated neuroticism score in MDD cases in our samples. Individuals with a diagnosis of MDD had significantly higher EPQN scores compared to healthy controls (all $p < 1.9. \times 10^{-279}$) in both GS:SFHS ($mean_{controls} = 3.16$; $mean_{cases} = 6.42$) and UKB ($mean_{controls} = 2.79$; $mean_{cases} = 5.64$). Neuroticism levels differ significantly between males and females. To control for this and any age/polygenic effects, which may account for differences in the prevalence of MDD, we created a matched set of cases and controls. The difference in neuroticism levels between cases and controls remained significant after matching the controls for PGC PRS_D, sex and age. (GS:SFHS: $mean_{controls} = 3.51$; UKB: $mean_{controls} = 2.97$; all $p < 2.7 \times 10^{-158}$; **Appendix A**: Supplementary Table 1).

2.4.1.1 Meta-analysis of stress-sensitivity in UKB and GS:SFHS

No SNPs were associated with stress-sensitivity at the genome-wide significant threshold ($p < 5 \times 10^{-8}$, **Figure 2.1**). However, 14 SNPs from 8 loci achieved suggestive p value ($p < 1 \times 10^{-5}$) ranging between $p = 8.9 \times 10^{-6}$ - 5.1×10^{-7} (Meta-analysis: **Table 2.1**; UKB and GS:SFHS: **Appendix A**: Supplementary Table 2 and Supplementary Table 3; Meta-analysis QQ-plot with λ: **Appendix A**: Supplementary Figure 2; UKB and GS:SFHS QQ-plots: **Appendix A**: Supplementary Figure 3). Traits with prior evidence of association with the nearest genes to the 8 lead SNPs were identified using dbGap and are shown in **Appendix A**: Supplementary Table 4. Comparison between the SNP association profile along the genome between stress-sensitivity GWIS and MDD GWAS meta-analyses is shown in Miami plots filtering for the most significant stress-sensitivity or MDD SNPs ($p < 0.001$; Meta-analysis: **Figure 2.2**; UKB and GS:SFHS: **Appendix A**: Supplementary Figure 4). No SNP with a p-value < 0.01 had a corresponding p-value in the alternate trait, suggesting that different variants contribute to depression and stress-sensitivity. Gene-based test identified *ZNF366* as the only gene achieving genome-wide significance ($p = 1.48 \times 10^{-7}$; Bonferroni-corrected

significance threshold $p < 2.79 \times 10^{-6}$; **Appendix A**: Supplementary Table 5 and Supplementary Figure 5). Using summary statistics from meta-analysis GWIS results, stress-sensitivity SNP-based heritability was estimated from LD score regression at 5.0% ($h^2 = 0.0499$, s.e. = 0.017, $p = 1.67 \times 10^{-3}$). Conversely, the SNP-based heritability for MDD and neuroticism were estimated at 9.6% ($h^2 = 0.0962$, s.e. = 0.0179, $p = 3.87 \times 10^{-8}$) and 10.1% ($h^2 = 0.1006$, s.e. = 0.0076, $p = 3.47 \times 10^{-40}$) respectively, using summary statistics from the meta-analysed GWAS of UKB and GS:SFHS.

LD score regression was performed to obtain genetic correlations between stress-sensitivity, MDD and neuroticism. As previously shown, there was a significant genetic correlation between MDD and neuroticism ($r_G = 0.637$, s.e. = 0.0704, $p = 1.39 \times 10^{-19}$). However, we found no evidence for a genetic correlation between stress-sensitivity and MDD (rG = -0.099, s.e. = 0.182, $p = 0.585$) or between stress-sensitivity and neuroticism ($r_G = 0.114$, s.e. = 0.107, $p = 0.285$).

2.4.1.2 Pathway enrichment, functional annotation and gene expression analyses

Lead SNPs from the GWIS meta-analysis were investigated using DEPICT. No gene showed statistically significant links to stress-sensitivity at a DEPICT false discovery rate (FDR) < 0.05. No significant result was found for either gene set analysis or tissue enrichment analysis at FDR < 0.05. Evidence of regulatory elements on normal cell lines/tissues was identified for 5 of the 12 lead SNPs (i.e. rs3762096, rs10987199, rs2221540, rs246565, rs319924). Two lead SNPs were associated with eQTLs: rs319924 (an intronic SNP in *EYS*) and rs9509508 (an intronic SNP in *LATS2*) and potentially regulate *LGSN/RP3-407E4.3* ($p = 6.31 \times 10^{-12}$/ $p = 1.15 \times 10^{-5}$) and *LATS2* ($p = 3.74 \times 10^{-8}$), respectively.

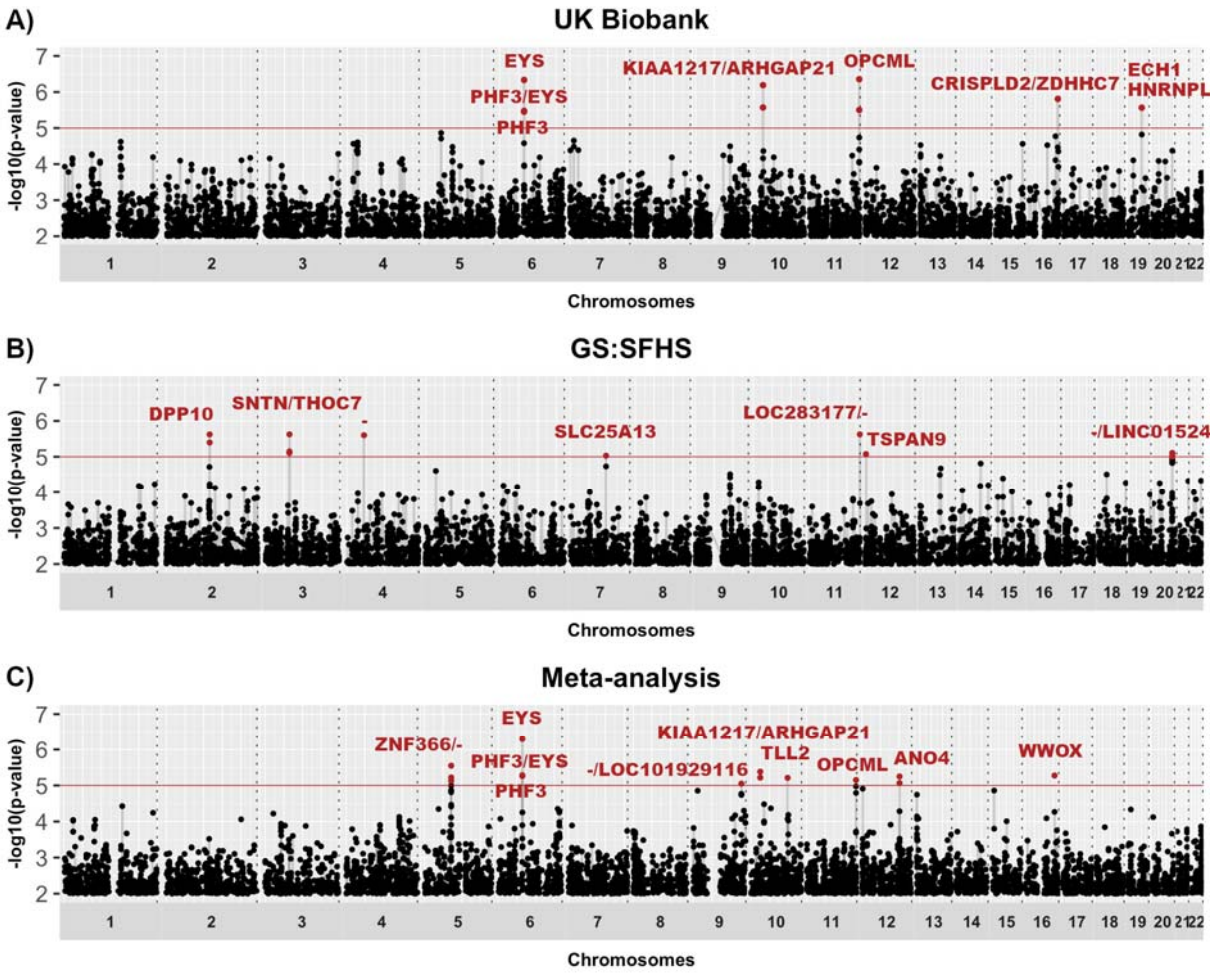


Figure 2.1 Manhattan plots showing stress-sensitivity associations. *Manhattan plots of the GWIS from (A) UKB, (B) GS:SFHS and (C) sample size weighted meta-analysis of UKB and GS:SFHS. The x-axis is chromosomal position and y-axis is the p-value (-log10 p-value) of association with stress-sensitivity effect. Suggestive genome-wide significance threshold ($p = 1 \times 10^{-5}$) is shown by solid line at y = 5. Genes or closest gene up- and down-stream from SNP position (/) are annotated. "-": No gene within 100kb of the SNP.*

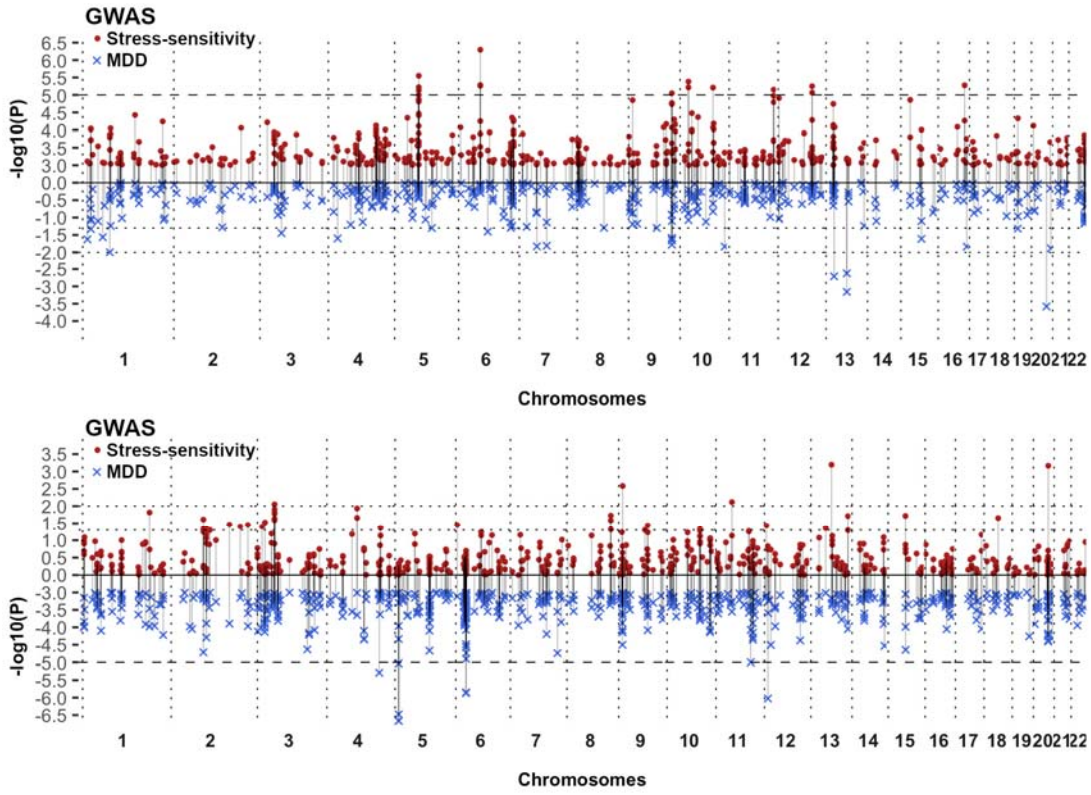


Figure 2.2 Miami plots showing comparison between association profile between stress-sensitivity GWIS and MDD GWAS. *Miami plots from meta-analysis filter at p = 1 x 10^{-3}: (A) filtering for stress-sensitivity p-values (•), (B) filtering for MDD p-values (×). The x-axis is chromosomal position and y-axis is the p-value (-log10 p-value) of association with stress-sensitivity (up; red dots) and MDD p-value (down; blue crosses). Dot line: genome-wide suggestive threshold (p = 1 x 10^{-5}) at the filtered effect; dashed lines: p = 0.01 and 0.05 at unfiltered effect.*

Table 2.1 Top 25 SNPs from meta-analysis of GWIS

Rank	CHR	SNP	BP	A1	Z-score	Effect[a]	*p*[b]	*p* (EPQN)[c]	*p* (MDD)[d]	GENE	POSITION[e]
1	6	rs319924	64487247	A	5.024	++	5.05×10^{-7}	0.376	0.637	*EYS*	Intronic
2	5	rs246565	71809247	A	-4.684	--	2.82×10^{-6}	0.248	0.589	*ZNF366*	5998bp 5'
3	10	rs2265265	24854876	A	4.604	++	4.15×10^{-6}	0.035	0.084	*KIAA1217 / ARHGAP21*	18104bp 3' / 17662bp 3'
4	6	rs1057530	64427095	A	-4.556	--	5.21×10^{-6}	0.636	0.840	*PHF3/EYS*	1677bp 3' / 2781bp 3'
5	16	rs7199110	78790765	A	-4.553	--	5.29×10^{-6}	0.661	0.741	*WWOX*	Intronic
6	6	rs10485358	64386060	A	-4.546	--	5.46×10^{-6}	0.390	0.902	*PHF3*	Intronic
7	12	rs10778077	101193988	A	4.54	++	5.62×10^{-6}	0.614	0.430	*ANO4*	Intronic
8	5	rs13358894	71803446	A	4.527	++	5.99×10^{-6}	0.257	0.651	*ZNF366*	197bp 5'
9	10	rs2256220	24856314	A	-4.524	--	6.06×10^{-6}	0.134	0.129	*KIAA1217 / ARHGAP21*	19542bp 3' / 16224bp 3'
10	10	rs3762096	98136250	A	-4.521	--	6.15×10^{-6}	0.437	0.149	*TLL2*	Intronic
11	11	rs2221540	132716369	A	-4.492	--	7.05×10^{-6}	0.468	0.364	*OPCML*	Intronic
12	5	rs10043659	71781839	A	4.483	++	7.37×10^{-6}	0.339	0.808	*ZNF366*	Intronic
13	12	rs10778078	101195088	A	-4.45	--	8.58×10^{-6}	0.599	0.456	*ANO4*	Intronic
14	9	rs10987199	128968987	A	-4.442	--	8.91×10^{-6}	0.199	0.026	*LOC101929116*	63416bp 3'
15	5	rs10042132	71789021	A	-4.416	--	1.01×10^{-5}	0.418	0.538	*ZNF366*	Intronic
16	11	rs10894606	132671611	A	-4.404	--	1.06×10^{-5}	0.438	0.587	*OPCML*	Intronic
17	12	rs7295089	2440464	A	4.372	++	1.23×10^{-5}	0.266	0.212	*CACNA1C*	Intronic
18	5	rs9293292	71696942	A	-4.351	--	1.36×10^{-5}	0.126	0.731	*PTCD2/ZNF366*	41762bp 3' / 42292bp 3'
19	15	rs3097437	27872136	A	4.346	++	1.38×10^{-5}	0.970	0.226	*GABRG3*	93762bp 3'
20	9	rs1999377	11919732	A	4.344	++	1.40×10^{-5}	0.436	0.064	-	Intragenic
21	5	rs6862221	71754962	A	4.342	++	1.41×10^{-5}	0.543	0.823	*ZNF366*	Intronic
22	5	rs9293289	71683885	A	-4.323	--	1.54×10^{-5}	0.395	0.510	*PTCD2/ZNF366*	28705bp 3' / 55349bp 3'
23	11	rs4575282	132719646	A	-4.313	--	1.61×10^{-5}	0.598	0.514	*OPCML*	Intronic
24	9	rs2417008	128970219	A	-4.3	--	1.71×10^{-5}	0.208	0.026	*LOC101929116*	62184bp 3'
25	9	rs7021461	128972210	A	4.299	++	1.72×10^{-5}	0.202	0.025	*LOC101929116*	60193bp 3'

[a]Effect direction in GS:SFHS and UK Biobank. [b,c,d]Significances of: [b]GWIS stress-sensitivity effect; [c]SNP main effect on neuroticism derived from GWAS meta-analysis of EPQN between UK Biobank and Generation Scotland; [d]SNP main effect on MDD derived from GWAS meta-analysis of MDD between UK Biobank and Generation Scotland. [e]Position of the SNP respect to closest gene transcripts within 100kb (including UTRs) from 5 prime (5') or 3prime (3').

2.4.1.3 Polygenic risk scores for stress-sensitivity predict MDD liability

PRS were used to investigate whether common variants affecting stress-sensitivity predict MDD risk. We generated PRS (PRS_{SS}) for stress-sensitivity based on the summary statistics from the GWIS. After 10,000 permutations, PRS_{SS} significantly predicted MDD risk in GS:SFHS using weights from the larger UKB summary data (Empirical-*p* = 0.04; $p = 5.2 \times 10^{-3}$; β = 0.078, s.e. = 0.028; best-fit *p* threshold = 0.005; **Appendix A**: Supplementary Table 6). On the liability scale, the MDD variance explained in GS:SFHS by PRS_{SS} was modest (R^2 = 0.195%). This was less than predicted by PRS weighted by the genetic main effects of MDD or neuroticism (PRS_D: R^2 = 0.368%; PRS_N: R^2 = 0.459%; **Table 2.2** and **Appendix A**: Supplementary Table 6). However, this association was not cross-validated in UKB using summary data from the smaller GS:SFHS GWIS (Empirical-*p* = 0.68; *p* = 0.23; β = 0.004, s.e. = 0.003; best-fit *p* threshold = 0.005; PRS_{SS} R^2 = 0.013%; **Appendix A**: Supplementary Table 6), likely due to lack of power as a result of the small discovery sample size. PRS_D (R^2 = 0.204%) and PRS_N (R^2 = 0.166%) derived from GS:SFHS significantly predicted MDD in UKB (**Table 2.2** and **Appendix A**: Supplementary Table 6).

Due to the known genetic correlations between MDD, neuroticism and stressful life events[288], models jointly fitting the effects of multiple PRS were analysed. Multiple regression analyses in GS:SFHS showed that, compared to PRS_D effects alone, the stress-sensitivity effect derived from the UKB GWIS effects significantly explains an additional 0.195% (a predictive improvement of 53.1%, $p = 5.1 \times 10^{-3}$; PRS_D: β = 0.112, s.e. = 0.029; PRS_{SS} : β = 0.078, s.e. = 0.028). The inclusion of PRS_{SS} in the full model, where PRS_{SS} was fitted along with both PRS_D and PRS_N weighted by GWAS summary statistics derived from UKB remained significant; explaining an additional 0.172% (a predictive improvement of 24.6%, $p = 8.5 \times 10^{-3}$; PRS_D: β = 0.093, s.e. = 0.029; PRS_N: β = 0.107, s.e. = 0.030; PRS_{SS}: β = 0.073, s.e. = 0.028). In models fitting PRS_D and PRS_N, the variances explained were non-additive, demonstrating the partial overlap between MDD risk prediction

from PRS_D and PRS_N main additive effects. This is consistent with the known genetic correlation between these two traits. An overlap was not seen between the variance explained by PRS_{SS} effect and the variance explained by PRS_D and/or PRS_N. Multiple regression analyses fitting PRS_D and PRS_N derived from worldwide consortiums (**Figure 2.3**) showed that the increased sample size from GWAS used to derive PRS_D resulted in an increment of MDD variance explained in GS:SFHS by PRS_D (from 0.368% to 1.378%). However, there was no change in the proportion of the variance explained by the PRS_{SS} in the full model (PRS_{SS} $p = 3.5 \times 10^{-3}$). These results suggest that PRS_{SS} explains a proportion of MDD risk not accounted for by PRS_D or PRS_N at current sample sizes. However, these findings were not cross-validated in UKB using PRS_{SS} derived from GS:SFHS GWIS, likely due to lack of power as a result of the small discovery sample size (**Appendix A**: Supplementary Figure 6).

2.4.1.4 Using stress-sensitivity to stratify MDD in GS:SFHS

MDD cases show significantly higher PRS_{SS} ($p = 2 \times 10^{-3}$) and PRS_D ($p = 1.8 \times 10^{-4}$) than controls. Association between MDD-related traits and stress-sensitivity risk quintiles was assessed on MDD cases in order to identify a subgroup of MDD patients, perhaps defining a characteristic aetiological subtype of MDD. However, stratification analysis failed, and no quintile based on PRS_{SS} nor its interaction with quintiles based on PRS_D showed statistically significant effects on any trait analyzed. Individuals with high PRS_{SS} were not significantly different from other cases for sex, MDD course, age of onset or episode count, nor neuroticism, mood disorder or schizotypal personality scores ($p > 0.05$; **Appendix A**: Supplementary Table 7). Results remained non significant when PRSs were fitted as continuous variables ($p > 0.05$).

Table 2.2 MDD risk prediction at best fits

UKB predicting on GS:SFHS

Weighted effect	Best fit threshold	# SNPs	R^2 (%)[d]	R^2 (%)[e]	*p*	Empirical-*p*
Stress-sensitivity	0.005	1,626	0.141	0.195	5.2×10^{-3}	0.0399
MDD[a]	0.1	22,771	0.265	0.368	1.3×10^{-4}	0.0015
EPQN[b]	0.4	65,276	0.330	0.459	1.8×10^{-5}	0.0002
MDD[a] + EPQN[b]	-	-	0.503	0.699	8.0×10^{-7}	-
joint models[c]	-	-	0.627	0.871	1.2×10^{-7}	-

PGC2 & GPC predicting on GS:SFHS

PGC2 MDD[a]	1	92,248	0.993	1.378	1.4×10^{-13}	≤0.0001
GPC EPQN[b]	0.01	3,521	0.108	0.149	0.014	0.1038
PGC2 MDD + GPC EPQN[b]	-	-	1.052	1.461	1.7×10^{-13}	-
joint models[c]	-	-	1.203	1.671	1.6×10^{-14}	-

GS:SFHS predicting on UKB

Weighted effect	Best fit threshold	# SNPs	R^2 (%)[a]	R^2 (%)[b]	*p*	Empirical-*p*
Stress-sensitivity	0.005	1,526	0.008	0.013	0.231	0.6841
MDD[a]	0.03	7,725	0.130	0.204	1.6×10^{-6}	≤0.0001
EPQN[b]	0.05	12,296	0.106	0.166	1.6×10^{-5}	0.0005
MDD[a] + EPQN[b]	-	-	0.197	0.309	2.8×10^{-8}	-
joint models[c]	-	-	0.206	0.322	6.6×10^{-8}	-

PGC2 & GPC predicting on UKB

PGC2 MDD[a]	0.5	64,113	0.919	1.440	3.4×10^{-37}	<0.0001
GPC EPQN[b]	0.03	8,761	0.066	0.104	6.5×10^{-4}	0.006
PGC2 MDD[a] + GPC EPQN[b]	-	-	0.950	1.488	2.9×10^{-37}	-
joint models[c]	-	-	0.958	1.501	1.5×10^{-36}	-

[a]major depressive disorder, [b]neuroticism score, [c]combined effect fitting all 3 PRS weighted by all the effects (i.e. stress-sensitivity, MDD and EPQN), [d]Nagelkerke's R^2 at observed scale, [e]R^2 on the liability scale.

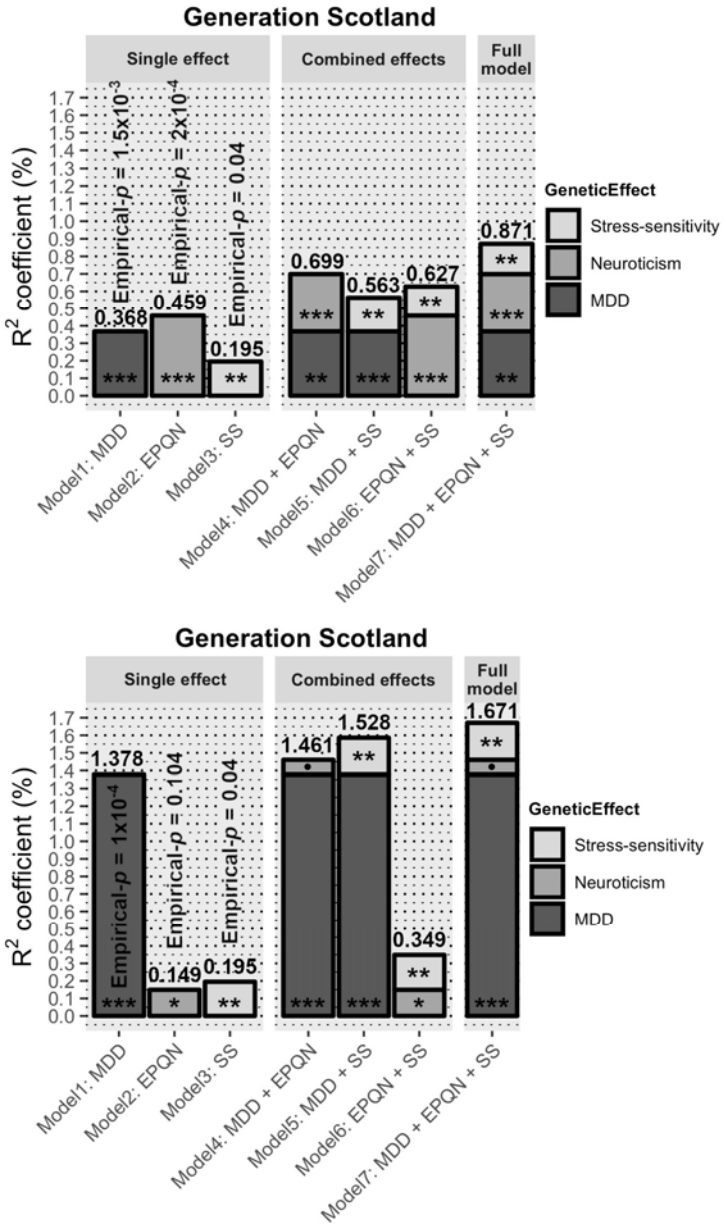


Figure 2.3 MDD is best predicted using multiple PRS. *MDD risk explained (R^2 coefficient (%); top bar values) on the liability scale by each PRS in GS:SFHS; weighted by GWAS main additive and GWIS stress-sensitivity effects independently and combined. (A) Using summary statistics from UKB as discovery sample. There is an increment on MDD risk prediction from adding PRS_{SS} to PRS_D model of 53.1% and 24.6% when combining PRS_{SS} with both MDD and neuroticism PRS. (B) Replication of fitting PRS_D and PRS_N using summary statistics from worldwide consortiums (i.e. PGC &GPC). Significance codes: p values *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.1; derived from likelihood ratio tests. SS stands for stress-sensitivity.*

2.5 Discussion

The existence of genetic variants affecting an individual's risk of depression in response to stress has been predicted previously[103,115,169] and is consistent with the departure from a simple additive genetic model seen in twin-studies of recurrent depressive disorder[82]. Through international research efforts such as the PGC and UK Biobank, there are ever-increasing sample sizes available for understanding the genetics of MDD. These resources are beginning, and will continue to, identify genome-wide significant loci[149,150,153]. However, the lack of environmental data and/or their reliability, makes the study of genetic individual's response to their negative effects, and their contribution to the onset of MDD and other stress-related disorders, difficult. As a way to address this limitation, we generated a proxy for stress-sensitivity through modelling the interaction between SNP allele and MDD status on neuroticism score in a GWIS approach. Thus, we sought to identify the genetic underpinnings of individual's sensitivity to stress response (stress-sensitivity) through those variants that contribute to higher neuroticism levels only in individuals with a lifetime diagnosis of MDD but not in healthy controls.

We performed a GWIS to identify loci showing differential effects on neuroticism scores in individuals with and without MDD (so called stress-sensitivity proxy). No SNPs reached genome-wide significance, but 14 SNPs from 8 loci reached suggestive significance levels (see **Appendix A**: Supplementary Table 4 for prior evidence of associated phenotypes). Enrichment analysis showed no evidence for enrichment of specific pathways or tissues. The top two loci, *PTP4A1-PHF3-EYS* and *ZNF366* have been previously associated with alcohol dependence[326-330], alcohol intake (dbGaP: phs000342) and glucocorticoid receptor function[331-333]. The most significant SNP in this study, rs319924, is an intronic variant in *EYS* that is a potential eQTL for *LGSN*[320], a gene previously associated with male-specific depression[334]. This is of particular interest given previous studies linking

alcohol consumption, stress and the risk of depression[335-340]. However, findings should be interpreted with caution, as these loci did not reach genome-wide significance at current sample size. Evidence of an eQTL effect was predicted for a lead SNP in *LATS2*, a positive regulator of histone methyltransferase activity[341] a process important in anxiety-related behaviours[342]. The prior association of the top two loci in this study with alcohol related-phenotypes suggests that genes involved in the sensitivity to stress may mediate the effects of stress on alcohol consumption. Some *PHF3* paralogs have been shown to be linked with depression and modulate stress response[343,344].

Gene-based analysis identified a genome-wide significant association between *ZNF366* and stress-sensitivity. *ZNF366* (also known as *DC-SCRIPT*) is a corepressor of transcription found in nuclear receptor complexes including the glucocorticoid receptor. *ZNF366* represses glucocorticoid receptor-mediated transcription in monocyte-derived dendritic cells[331]; and may act through histone deacetylases to modulate immune response[332]. There is evidence from a large-scale mRNA display study that *PHF3*, in the region underlying the most significant peak in the single SNP analysis, may also interact, directly or indirectly, with the glucocorticoid receptor (IntAct database [333]) but this has not been confirmed. These results reinforce the hypothesis that our proxy for stress-sensitivity truly reflects the genetic architecture of sensitivity to respond to stress.

We estimated a significant lower bound on common SNP-based heritability for stress-sensitivity of 5%. Whilst the known genetic overlap between MDD and neuroticism was detectable, the lack of genetic correlation with stress-sensitivity, reinforced by results from multiple regression analyses, indicated a lack of significant overlap in the genetics factors underpinning stress-sensitivity and MDD or neuroticism. This analysis may be limited by our sample size, although using the largest available meta-analyses of MDD and neuroticism[150,321] did not decrease the proportion of liability explained by the PRS_{SS}. We note, that as such meta-analyses increase in size it is likely, as

with the effects of smoking in schizophrenia[345,346], that the indirect genetic effects of the environment on the risk of depression will be detected by GWAS. However, through studies such as ours, or similar, the mechanism for the effect of the risk alleles may be clarified.

Further, we show that such genetic information in stress-sensitivity could significantly improve the proportion of liability to MDD predicted by PRS based only on additive genetic effects on MDD identified by large GWAS. The summary results from the GWIS were used to derive a PRS reflecting the genetic difference in stress-sensitivity. This variable significantly predicted liability to MDD in GS:SFHS (p = 5.2 x 10^{-3}, Empirical-p = 0.04 after 10,000 permutations), although this finding could not be replicated in UKB (Empirical-p = 0.68), likely due to lack of power. This is consistent with the expectation that the larger the discovery sample (i.e. UKB), the greater the accuracy of the weighting and the more predictive the PRS[170]. Multiple regression models in GS:SFHS suggest that inclusion of PRS weighted by stress-sensitivity significantly improves MDD prediction over use of either MDD and/or neuroticism weighted PRS alone (improvement in full model p = 8.5 x 10^{-3}). However, we were unable to identify a subgroup of MDD cases with higher PRS_{SS}. The polygenic interaction approach used in our study may, therefore, improve the interpretation of both positive and negative findings from GWAS studies (i.e. pathways and mechanisms involved, lack of replication, or negative findings in variants mediating environmental effects). Added to paralleling recent developments in GWAS analyses, it may maximize our power to detect gene-by-environment effects in this heterogeneous disorder.

Future studies will be required to further investigate the effects of adverse life events in individuals with high or low polygenic risk scores for stress-sensitivity. However, the methodology presented allows addressing the genetic response to negative outcomes via proxy in the absence of prospective environmental data.

Here we identify an independent set of risk variants for an individual's response to negative outcomes and show that incorporating information across many loci provides clear and replicable evidence for a genetic effect of stress-sensitivity on MDD risk; identifying a potential genetic link with alcohol intake. These results require further study, but may inform treatment of comorbid alcohol dependency and depression.

Chapter 3 Enrichment of genetic variation conferring MDD risk in glucocorticoid-related genesets: partitioning risk based on main additive contributions to MDD, neuroticism and the proxy for stress-sensitivity

3.1 Introduction

In **chapter 2**, the strongest signals for the proxy of stress-sensitivity coming from genetic variants associated with the glucocorticoid receptor function (e.g. *ZNF366* and the *PTP4A1-PHF3-EYS* locus). Thus, within this chapter, I aim to investigate the enrichment of genetic contributions to risk of MDD within sets of genes related to the glucocorticoid signalling and stress response pathways.

In response to acute periods of psychological stress, a cascade of adaptive physiological changes occurs through the hypothalamic-pituitary-adrenal (HPA) axis involving the brain, the pituitary gland and the adrenal glands[347]. This is a complex neuroendocrine pathway responsible for the regulation and eventually release of glucocorticoid hormones, cortisol in humans, from the adrenal cortex to the bloodstream. The relationship between psychological stress, HPA axis function and major depressive disorder (MDD) has been widely studied but the mechanisms underlying this linkage remain unclear[72,348,349]. In response to stress, glucocorticoid hormones regulate the HPA axis activation through negative feedback, acting in multiple target tissues including the brain where they effect the structure and plasticity of neurons[350,351]. While the signalling cascade is activated, non-essential functions such as controlling metabolism or immune response are inhibited to prioritize the most essential ones, such as behavioural arousal and cardiovascular activity[349,352], partially explaining why chronic stress during long times of period is such a problem for health. For example, cortisol

reduces histamine and proinflammatory cytokine secretion resulting in anti-inflammatory effects[353]. Conversely, resistance to glucocorticoid hormones is reflected in hyperactivity of the HPA axis that can activate the immune system and inflammatory processes, a characteristic feature of depression[354]. Hyperactivity of the HPA axis and inflammatory process in adult depression as response to stress may be part of the same pathophysiological process[72]. Elevated cortisol also cause insulin resistance, which may explain why conditions such Type II diabetes and metabolic syndrome present higher prevalence in recurrent depression[355]. Therefore, psychosocial stress could induce strong effects on the HPA axis. Stress-induced HPA axis dysfunction is associated with depressive symptoms and is frequently found in individuals with MDD[356-359]. Permanent changes in the HPA axis caused by early-life stress have been demonstrated[360,361] and could be a major trigger for the development of depressive disorders in those individuals with genetic vulnerability. Such dysregulation of the HPA axis also contributes to craving, alcohol dependence, damage in the brain's reward system, and their associated problematic behaviours[362]. The association seen in **chapter 2** between the top 2 loci detected in the genome-wide interaction study (GWIS) of a proxy for stress-sensitivity and alcohol related conditions suggested that genes involved in sensitivity to stress may mediate the association between stress and alcohol consumption. In line, the interplay between a dysregulation of the HPA axis and exposure to stress has been reported to also increase the risk of developing alcohol-related disorders[362,363].

Such proxy for stress-sensitivity modeled in **chapter 2** was estimated from the MDD-dependent increment in neuroticism score seen in individuals with MDD compared to healthy controls. Although neuroticism is report to be, at population level, a stable trait strongly determined by genetics, there is evidence that individuals experience changes in neuroticism throughout their lives[217-220,299]. Such change is attributed to the effects of environmental factors[217-220], however, genetic factors may moderate the association between stress and such increment in neuroticism. Increments of neuroticism sustained over time have been associated with stressful life events[220]. Genetic contributions to the MDD-depended change in neuroticism, here so-

called stress-sensitivity, may differ from the strong genetic contributions to the stable component of neuroticism.

At the molecular level, glucocorticoids diffuse across cell membranes and bind cytoplasmic glucocorticoid receptors, which are then translocated into the nucleus. In the nucleus these receptors bind to specific glucocorticoid receptor binding sites acting as transcription factors through specific glucocorticoid response elements within the promoter regions of glucocorticoid-regulated genes, which consist of a short DNA sequence that mediates the effect of glucocorticoids on gene expression[364,365]. Such glucocorticoid response elements are found in many genes and can both activate or inhibit (i.e. negative glucocorticoid response elements) gene expression[364-366]. The glucocorticoid receptor binding sites have been characterized in humans and mouse in previous studies, mainly in blood, using protein-binding microarrays that combine Chromatin ImmunoPrecipitation with next generation sequencing (i.e. ChipSeq techniques), using dexamethasone as an analog of glucocorticoid hormones[367-369]. Polman *et al.*[368] compared the degree of overlap of glucocorticoid receptor binding sites in neuronal PC12 cells from rat with glucocorticoid receptor binding sites identified in human lung carcinoma (A549) in Reddy *et al.*[367], and in mouse adipocytes (3T3-L1) from a third study[369]. The majority of glucocorticoid receptor binding sites identified in PC12 cells (87%) appeared unique to neuronal PC12 cells. Only 7% of them were shared with A549 cells and 11% with 3T3-L1 cells, with only 4% of glucocorticoid receptor binding sites overlapping across all three cell types. This low degree of overlap in glucocorticoid binding sites across cell types concurred with other ChIP-Seq studies[370].

Sensitivity to long-term effects of psychological stress in the development of MDD may be modulated by genetic variability. Genes associated with the correct functioning of the HPA axis such as *CRHR1*, *FKBP5* and *NR3C2*[371-374] have been reported to moderate the effects of stress in depression and amygdala function[375,376]. Arloth *et al.*[377] suggested that genetic differences in the activation of gene expression induced by glucocorticoid receptors might mediate MDD risk by altering a network of co-expressed stress-sensitive

genes that are responsive to glucocorticoids in the brain. The glucocorticoid stress response pathway may be crucial for susceptibility to MDD, or susceptibility to environmental stress leading to depression. I hypothesize that there will be a significant enrichment in SNPs associated with the proxy for stress-sensitivity in the glucocorticoid stress response pathway.

In this chapter, I use gene ontologies to select genes ontologically related with the HPA axis to define a geneset representative of “up-stream” glucocorticoid signalling. I used freely available data on glucocorticoid receptor binding sites[367,368] to define two sets of genes representative of “down-stream” glucocorticoid response including genes selected as candidate glucocorticoid-regulated targets through activation of glucocorticoid response elements (i.e. genes of which transcription profile is likely to be modulated by glucocorticoid receptors) in two different cell types: human lung epithelial carcinoma[367] and pheochromoytoma of the rat adrenal medulla[368].

The aim of this chapter is to investigate the contribution of genetic variation to risk to MDD from genetic variation in glucocorticoid-related genes. I assess whether the genetic contributions to both MDD, neuroticism and sensitivity to stress that increase the risk to liability are significantly more clustered in the three glucocorticoid-related genesets than expected by random chance. Enrichment of common variants conferring risk of MDD in the “down-stream” sets of glucocorticoid response genes was present, including an enrichment of contributions from stress-sensitivity to MDD when additive main effects of MDD and neuroticism are taken into account. The results suggest that characterizing genes modulated by glucocorticoid receptors on neuronal cell lines may benefit the assessment of the link between glucocorticoid receptor-mediated gene expression and risk of depression.

3.2 Materials and methods

3.2.1 Cohorts' profiles

This study uses genotype data from Generation Scotland (GS) and UK Biobank (UKB) from unrelated individuals previously described in **chapter 2**. Further details about study ethics, cohort descriptions and phenotype assessment are described previously. MDD status and the Eysenck Personality Questionnaire-Revised Short Form's Neuroticism Scale were available in both UKB and GS cohorts. Briefly, dataset used from UKB contains genome-wide genotyping data available on 109,282 unrelated individuals (KING-estimated kinship coefficient < 0.0442) aged 40-79 with neuroticism data available, of which 23,092 had MDD data after the quality control process. Dataset from GS includes genome-wide genotyping data on 7,233 unrelated individuals (pi-hat < 0.05) aged 18-92 of which: 7,190 had clinical data on MDD; 7,196 individuals had data on neuroticism; and 7,155 had data on both neuroticism and MDD, after the quality control process. See **chapter 2** for further details (e.g. genotyping, imputation and quality control)[378].

3.2.2 Glucocorticoid-related genesets design

Three genesets were defined targeting the glucocorticoid stress response pathway. The list of genes forming each geneset can be found in **Appendix B**: Supplementary Table 1. All gene are selected using the following criteria: known genes, as well as genes of unknown function with identifiers: LOC# (# = 6 digits number), KIAA@ (@ = 4 digits) and homo sapiens chromosome "x" open reading frame "y" ($C_x orf_y$); long non-coding RNA and long intergenic non-coding RNA genes with identifiers AC# and LINC#. All other classes of gene elements or identifiers were excluded from the geneset design. All gene products were obtained from GENCODE (UCSC Genome Browser: http://genome.ucsc.edu) assembly Dec. 2013 (GRCh38/hg38) but gene coordinates were remapped to assembly Feb. 2009 (GRCh37/hg19) using the Batch Coordinate Conversion (liftOver) tool from UCSC Genome Browser to match the GS genotype data. HGNC symbols are used to define the genes, with -50 kb upstream and +50 kb downstream used to define the gene

boundaries when extracting SNPs from the genotype data (since LD blocks are likely to be contained within these intervals) using python scripts of own design. The list of genes without SNP data in GS is presented in **Appendix B**: Supplementary Table 2. Human orthologous search and ID conversions were carried out using Ensembl[379] and BioMart[380] when required. To assess each geneset design, I investigated gene ontology terms enriched in each geneset in comparison to the full list of genes in the human genome (GRCh37/hg19; p-value threshold = 1 x 10^{-5}) using Gorilla[381]; gene ontology terms divided in biological processes, molecular functions and cellular components. Revigo[382] was used to visualize summary results. In geneset 3, the *p*-value threshold was increased to 1 x 10^{-4} to see the most enriched ontologies for molecular functions. Hypergemoetric test was applied to test whether the degree of overlap between gensets was significantly higher than expected by chance.

3.2.2.1 Geneset 1: genes ontologically related with HPA axis

Geneset 1 contains genes ontologically-related to the glucocorticoid signalling. These were extracted from the web application AmiGO[383] from the Gene Ontology (GO) website using as key words: glucocorticoid, mineralocorticoid, cortisol, CRH (corticotropin-releasing hormone), CRF (corticotropin-releasing factor), ACTH (adrenocorticotropic hormone) and POMC (proopiomelanocortin). All gene products non-human genes were converted to human orthologous genes. There is no SNP data for 8 of those genes in GS (**Appendix B**: Supplementary Table 2). This geneset contains 348 genes (7 from chromosome X).

3.2.2.2 Geneset 2: genes overlapping a glucocorticoid receptor binding sequence or within 50Kb from an A549 cell line (human lung epithelial carcinoma)

Geneset 2 represents all genes overlapping a genomic binding site of glucocorticoid receptor, or the closest gene to that sequence within 50 kb (when there is no overlap) from an immortalized A549 cell line derived from human lung epithelial carcinoma (from a 58-year-old Caucasian male). The genomic binding sites for glucocorticoid receptor were previously defined by Reddy *et al.*[367] using Chromatin ImmunoPrecipitation combined with next generation sequencing (ChIP-Seq) followed by next-generation DNA sequencing to measure glucocorticoid receptor binding in response to 100nM

of dexamethasone, a synthetic glucocorticoid used primarily in the treatment of inflammatory disorders. Data manipulation and gene extraction was carried out using BEDtools[384] and Galaxy server (http://galaxyproject.org/). This geneset contains 5,936 genes (163 from chromosome X).

3.2.2.3 Geneset 3: genes overlapping a glucocorticoid receptor binding sequence or within 50Kb from a PC12 cell line (pheochromoytoma of the rat adrenal medulla).

Geneset 3 contains human orthologous genes overlapping a genomic binding site of glucocorticoid receptor, or the closest gene in a range of 50 kb when there is no-overlap, from a PC12 cell line (cultured and differentiated into a neuronal phenotype for ten days) derived from a pheochromocytoma of the rat adrenal medulla using ChIP-seq. Human orthologous genes were obtained from the list of rat genes that met these requirements from a previous study by Polman *et al*.[368]. Approximately one third (31%) of all significant genomic binding sites were located within a gene. Human orthologous search from rat genes and ID conversions were carried out using Ensembl and BioMart. This geneset contains 633 genes (13 from chromosome X).

3.2.3 Polygenic profiling

Polygenic risk scores (PRS) were produced using PRSice[313] and standardized to mean of 0 and standard deviation of 1. PRS were weighted using GWAS and GWIS summary statistics for MDD, neuroticism, and stress-sensitivity conducted in UKB and GS and previously used to assess MDD risk at genome-wide level (discovery sample size: UKB GWAS of MDD = 23,092; UKB GWAS of neuroticism = 109,282; UKB GWIS of stress-sensitivity = 23,092; GS GWAS of MDD = 7,190; GS GWAS of neuroticism = 7,196; and GS GWIS of stress-sensitivity = 7,155). Further details about GWAS and GWIS were reported in **chapter 2**. First, we clumped all SNPs present in both the discovery and target samples (GS and UKB) to obtain a set of independent SNPs in approximate linkage equilibrium ($r2 < 0.1$, within a 250kb window). From the clumped SNPs, which are used at the genome-wide level prediction, a subset was selected (see “Glucocorticoid-related genesets design” above) for each of the glucocorticoid-related genesets. Thus, each geneset only contains as predictors independent SNPs from the genome-wide prediction, making each geneset prediction proportional to its contribution to

the genome-wide set, if the genome-wide MDD risk was polygenic and randomly distributed along the genome. The number of SNPs selected for each dataset and its contribution to the genome-wide set are shown in **Table 3.1**. Finally, three sets of PRS were generated for each dataset in each cohort (GS and UKB) using the summary results generated in the previous **chapter 2**: the genetic effects of MDD derived from GWAS of MDD (PRS_{MDD}), the genetic effects of neuroticism derived from GWAS of neuroticism (PRS_N), and the genetic effects of stress-sensitivity derived from GWIS (PRS_{SS}).

Table 3.1 Density of independent SNPs in each data set

Set	Genome-wide	Geneset 1	Geneset 2	Geneset 3
Target sample		**UK Biobank**		
Number of SNP	98,556	2,266	39,032	5,640
Genome-wide representation	100 (%)	2.30 (%)	39.60 (%)	5.72 (%)
Target sample		**Generation Scotland**		
Number of SNP	97,678	2,245	38,623	5,596
Genome-wide representation	100 (%)	2.30 (%)	39.54 (%)	5.73 (%)

All SNPs are independent in approximate linkage equilibrium (r2 < 0.1, within a 250kb window). Genome-wide representation is the ratio between the number of SNPs in a set and the total number of genome-wide SNPs.

3.2.4 PRS prediction of MDD risk

Using UKB (the largest cohort) summary statistics as the discovery sample to weight PRS and logistic regression models, we estimated the variance of MDD risk explained by both PRS_D (model 1), PRS_N (model 2) and PRS_{SS} (model 3) independently in GS at genome-wide level. Then, models fitting additively both PRS_D and PRS_N (model 4), and all three PRS together (model 5) were tested. A null model was also estimated from the direct effects of all covariates on MDD (null model). All models were adjusted by sex, age and 20 PCs.

model 1: $MDD \sim PRS_D + Covariates$

model 2: $MDD \sim PRS_N + Covariates$

model 3: $MDD \sim PRS_{SS} + Covariates$

model 4: $MDD \sim PRS_D + PRS_N + Covariates$

model 5: $MDD \sim PRS_D + PRS_N + PRS_{SS} + Covariates$

null model: $MDD \sim Covariates$

All predictions and models were repeated and implemented partitioning by genesets. Likelihood ratio tests were used to assess: (i) the significance of each PRS (models 1, 2 and 3 against null model), (ii) the significance of the full risk model (model 5 against null model) and (iii) the significance of PRS_{SS} effect on the residual variance of MDD after considering the effects of PRS_D and $PRSS_N$ (model 5 against model 4). Nagelkerke's R^2 coefficients were estimated to quantify the proportion of MDD liability explained at the observed scale. Generalized linear models were implemented in R 3.1.3[322]. Cross-validation was conducted in UKB (target sample) using GS as the discovery sample. In UK Biobank, MDD was adjusted for centre, array and batch as random effects and the residuals scaled and used for prediction. This resulted in a quasi-binomial distribution of MDD status in UKB. Thus, quasi-binomial regressions were implemented. PRS predictive models were adjusted for age, sex and 15 principal components (PCs; UKB Data Dictionary items #22009.01 to #22009.15).

3.2.5 Enrichment of MDD risk

The expected amount of MDD risk variance that is explained by each geneset was estimated as that expected if genome-wide MDD risk is polygenic and homogeneously distributed along the genome using the genome-wide estimations of MDD risk reported in **chapter 2**. The ratio between the observed MDD risk explained by each genetic contribution and that expected was used as indicator of enrichment in each geneset. To assess significance of MDD risk enrichment on each geneset, we derived a null predictive distribution by permuting 1,000 times MDD phenotype from models 1, 2 and 3. From each permutation and for each set of PRS, PRS at the best *p*-threshold predicting the permuted phenotype were selected and combined into models 4 and 5 respectively to predict real MDD status. *P*-values from

likelihood ratio test applied at each permutation testing models 1, 2 and 3 against null model, (ii) model 5 against null model, and (iii) model 5 against model 4 were used to create an empirical cumulative distribution function to adjust the observed *p*-values. The significance threshold of empirical *p*-values was fixed at $\alpha = 0.05$.

3.3 Results

3.3.1 Genesets overlap

Top enriched gene ontology terms in predefined genesets compared to the full list of genes in the human genome are shown in **Tables 3.2, 3.3** and **3.4**. Visualization of summary results is shown in **Appendix B**: Supplementary 1-9.

The degree of overlap of genes across genesets was low (**Figure 3.1**), but significantly higher than expected by chance in comparison with a genomic background of 26,292 genes (Human Genome version 19; overlap genesets 1-2: *p*-value = 6.13 x 10^{-17}, geneset 1-3: *p*-value = 8.59 x 10^{-4}, genesets 2-3: *p*-value = 2.36 x 10^{-51}). 49% of genes from geneset 3 (from pheochromoytoma of the rat adrenal medulla) were shared with geneset 2 (from human lung epithelial carcinoma'), and 2.84% with geneset 1 (genes ontologically related with HPA axis). 57% of genes in geneset 1 were unique to geneset 1, 92% of genes in geneset 2 were unique to geneset 2, and 50% of genes in geneset 3 were unique to geneset 3.

Table 3.2 The functional analysis of genes from geneset 1
Top 5 gene ontology terms.

Category	GO ID	Name	Count	*p*-value	FDR q-value
BP	GO:0051384	response to glucocorticoid	136	5.93×10^{-240}	8.07×10^{-236}
BP	GO:0031960	response to corticosteroid	139	1.11×10^{-239}	7.59×10^{-236}
BP	GO:0048545	response to steroid hormone	151	2.28×10^{-176}	1.04×10^{-172}
BP	GO:0009725	response to hormone	176	2.73×10^{-165}	9.28×10^{-162}
BP	GO:0033993	response to lipid	165	1.03×10^{-144}	2.79×10^{-141}
MF	GO:0005102	receptor binding	114	6.44×10^{-45}	2.73×10^{-41}
MF	GO:0001664	G-protein coupled receptor binding	36	1.62×10^{-22}	3.42×10^{-19}
MF	GO:0051427	hormone receptor binding	30	4.02×10^{-21}	5.68×10^{-18}
MF	GO:0005126	cytokine receptor binding	31	1.52×10^{-16}	1.61×10^{-13}
MF	GO:0035259	glucocorticoid receptor binding	9	2.29×10^{-16}	1.94×10^{-13}
CC	GO:0005615	extracellular space	93	4.84×10^{-34}	7.82×10^{-31}
CC	GO:0044421	extracellular region part	150	2.91×10^{-24}	2.35×10^{-21}
CC	GO:0005576	extracellular region	79	7.33×10^{-20}	3.95×10^{-17}
CC	GO:0044444	cytoplasmic part	213	8.91×10^{-16}	3.60×10^{-13}
CC	GO:0097458	neuron part	63	6.33×10^{-15}	2.05×10^{-12}

BP biological process, *MF* molecular function, *CC* cellular component, *Count* is the number of genes from the geneset.

Table 3.3 The functional analysis of genes from geneset 2

Top 5 gene ontology terms.

Category	GO ID	Name	Count	*p*-value	FDR q-value
BP	GO:0044763	single-organism cellular process	2791	5.30×10^{-39}	7.22×10^{-35}
BP	GO:0048518	positive regulation of biological process	1670	1.61×10^{-36}	1.09×10^{-32}
BP	GO:0044699	single-organism process	3236	3.99×10^{-35}	1.81×10^{-31}
BP	GO:0010646	regulation of cell communication	1007	9.30×10^{-32}	3.16×10^{-28}
BP	GO:0023051	regulation of signalling	963	2.21×10^{-31}	6.01×10^{-28}
MF	GO:0019899	enzyme binding	570	4.52×10^{-20}	1.92×10^{-16}
MF	GO:0008092	cytoskeletal protein binding	312	3.72×10^{-18}	7.88×10^{-15}
MF	GO:0005515	protein binding	2902	7.01×10^{-18}	9.92×10^{-15}
MF	GO:0003779	actin binding	165	1.54×10^{-13}	1.64×10^{-10}
MF	GO:0043168	anion binding	816	2.09×10^{-13}	1.78×10^{-10}
CC	GO:0044444	cytoplasmic part	2221	2.17×10^{-20}	3.50×10^{-17}
CC	GO:0005737	cytoplasm	1452	8.21×10^{-18}	6.63×10^{-15}
CC	GO:0030054	cell junction	406	3.77×10^{-17}	2.03×10^{-14}
CC	GO:0070161	anchoring junction	205	6.18×10^{-17}	2.50×10^{-14}
CC	GO:0005912	adherens junction	198	1.31×10^{-16}	4.24×10^{-14}

BP biological process, *MF* molecular function, *CC* cellular component, *Count* is the number of genes from the geneset.

Table 3.4 The functional analysis of genes from geneset 3

Top 5 gene ontology terms.

Category	GO ID	Name	Count	*p*-value	FDR q-value
BP	GO:0048522	positive regulation of cellular process	215	1.61×10^{-10}	2.19×10^{-6}
BP	GO:0044699	single-organism process	437	1.47×10^{-9}	9.98×10^{-6}
BP	GO:0048518	positive regulation of biological process	236	2.14×10^{-9}	9.69×10^{-6}
BP	GO:0044763	single-organism cellular process	375	1.06×10^{-8}	3.62×10^{-5}
BP	GO:0032502	developmental process	212	1.84×10^{-8}	5.01×10^{-5}
MF	GO:0004672	protein kinase activity	39	2.07×10^{-5}	8.77×10^{-2}
MF	GO:0043168	anion binding	119	5.76×10^{-5}	1.22×10^{-1}
MF	GO:0004683	calmodulin-dependent protein kinase activity	7	6.55×10^{-5}	9.26×10^{-2}
MF	GO:0016301	kinase activity	45	8.01×10^{-5}	8.50×10^{-2}
MF	GO:0016773	phosphotransferase activity, alcohol group as acceptor	42	8.66×10^{-5}	7.35×10^{-2}
CC	GO:0030054	cell junction	70	8.81×10^{-8}	1.44×10^{-4}
CC	GO:0044456	synapse part	42	1.52×10^{-7}	1.24×10^{-4}
CC	GO:0097458	neuron part	67	4.42×10^{-6}	2.40×10^{-3}
CC	GO:0044444	cytoplasmic part	302	8.63×10^{-6}	3.52×10^{-3}
CC	GO:0097060	synaptic membrane	23	9.06×10^{-6}	2.95×10^{-3}

BP biological process, *MF* molecular function, *CC* cellular component, *Count* is the number of genes from the geneset.

Overlapping genes with geneset 1 are highlighted in **Appendix B**: Supplementary Table 1. Just 16 genes (*ARNTL*, *ASS1*, *C1QTNF1*, *CDKN1A*,

CRY2, *DKK3*, *DUSP1*, *GAL*, *IL6R*, *KLF9*, *NEFL*, *PER1*, *PIK3R1*, *SFTPB*, *SLIT3*, *STAT5B*) were shared within all three glucocorticoid-related genesets (**Figure 3.1**). Gene ontology terms enriched in this set of overlapping genes were sought. Enriched biological processes are shown in **Table 3.5** and **Appendix B**: Supplementary Figure 10. There was no gene ontology term enriched in molecular functions or cellular component categories (*p*-threshold = 1 x 10^{-5}). The top biological process was that of rhythmic process that contained seven genes including *PER1* (Period Circadian Regulator 1), the primary circadian pacemaker in the mammalian brain. Genes in the PER family encode components of the circadian rhythms of locomotor activity, metabolism, and behaviour. PER1 is upregulated by CLOCK/ARNTL heterodimers but then represses this upregulation in a feedback loop using PER/CRY heterodimers to interact with CLOCK/ARNTL[385]. Notably, ARNTL and CRY2, also fall within this set of overlapping genes.

Figure 3.1 Venn diagram. *Representation of overlapping genes between glucocorticoid-related genesets.*

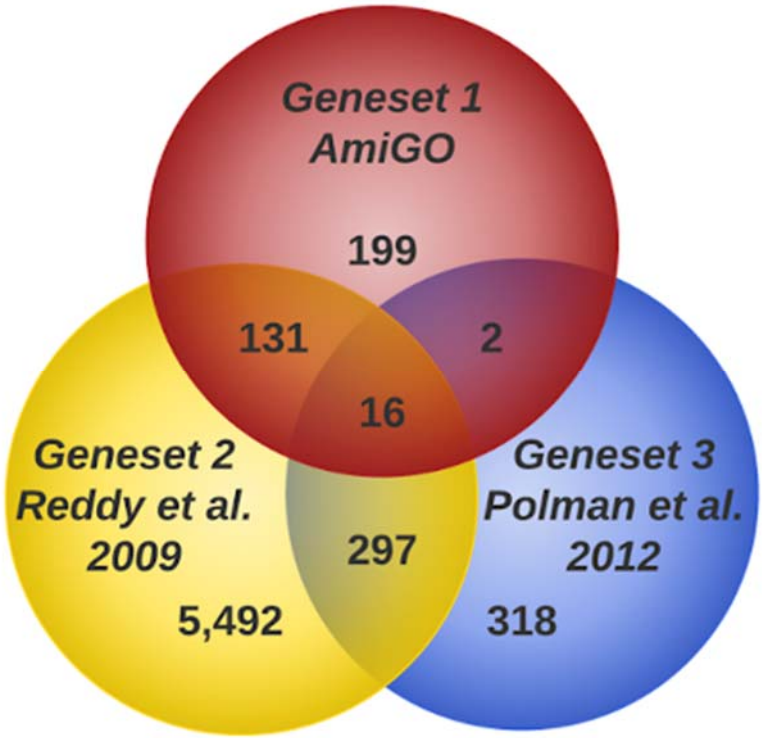


Table 3.5 The functional analysis of 16 genes overlapping all three genesets (biological processes)

GO ID	Name	*p*-value	FDR q-value	Genes
GO:0048511	rhythmic process	$5.13x10^{-10}$	$8.00x10^{-6}$	ARNTL, SLIT3, CRY2, ASS1, STAT5B, PER1 and KLF9.
GO:0009725	response to hormone	$5.46x10^{-10}$	$4.25x10^{-6}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, ASS1, STAT5B, KLF9 and PIK3R1
GO:0051384	response to glucocorticoid	$1.10x10^{-9}$	$5.72x10^{-6}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1 and KLF9
GO:0031960	response to corticosteroid	$2.16x10^{-9}$	$8.41x10^{-6}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1 and KLF9
GO:1901654	response to ketone	$5.31x10^{-9}$	$1.65x10^{-5}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1 and KLF9
GO:2000323	negative regulation of glucocorticoid receptor signalling pathway	$9.05x10^{-9}$	$2.35x10^{-5}$	ARNTL, CRY2 and PER1
GO:1901700	response to oxygen-containing compound	$1.15x10^{-8}$	$2.55x10^{-5}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, STAT5B, ASS1, PER1, KLF9 and PIK3R1
GO:0048545	response to steroid hormone	$1.22x10^{-8}$	$2.38x10^{-5}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1 and KLF9
GO:0009719	response to endogenous stimulus	$2.48x10^{-8}$	$4.30x10^{-5}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, STAT5B, ASS1, KLF9 and PIK3R1
GO:2000322	regulation of glucocorticoid receptor signalling pathway	$2.53x10^{-8}$	$3.94x10^{-5}$	ARNTL, CRY2 and PER1
GO:0014070	response to organic cyclic compound	$4.24x10^{-8}$	$6.01x10^{-5}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1, STAT5B, PER1 and KLF9
GO:0007623	circadian rhythm	$6.29x10^{-8}$	$8.17x10^{-5}$	ARNTL, CRY2, ASS1, PER1 and KLF9
GO:0010033	response to organic substance	$6.35x10^{-8}$	$7.62x10^{-5}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, STAT5B, ASS1, PER1, IL6R, KLF9 and PIK3R1
GO:0048523	negative regulation of cellular process	$1.45x10^{-7}$	$1.62x10^{-5}$	ARNTL, C1QTNF1, SLIT3, NEFL, CDKN1A, CRY2, STAT5B, ASS1, PER1, PIK3R1, DKK3, GAL, DUSP1 and KLF9
GO:0032870	cellular response to hormone stimulus	$1.79x10^{-7}$	$1.86x10^{-5}$	DUSP1, SLIT3, ASS1, STAT5B, PIK3R1 and KLF9
GO:0042221	response to chemical	$6.29x10^{-7}$	$6.13x10^{-5}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, STAT5B, ASS1, PER1, IL6R, KLF9 and PIK3R1
GO:0033993	response to lipid	$9.77x10^{-7}$	$8.96x10^{-5}$	DUSP1, SLIT3, NEFL, CDKN1A, ASS1, STAT5B and KLF9
GO:0042493	response to drug	$1.03x10^{-6}$	$8.93x10^{-5}$	GAL, DUSP1, SLIT3, NEFL, CDKN1A, ASS1 and KLF9
GO:0042634	regulation of hair cycle	$1.31x10^{-6}$	$1.07x10^{-5}$	ARNTL, GAL and PER1
GO:0048519	negative regulation of biological process	$1.39x10^{-6}$	$1.08x10^{-5}$	ARNTL, C1QTNF1, SLIT3, NEFL, CDKN1A, CRY2, STAT5B, ASS1, PER1, PIK3R1, DKK3, GAL, DUSP1 and KLF9
GO:0010243	response to organonitrogen compound	$1.59x10^{-6}$	$1.18x10^{-5}$	GAL, DUSP1, NEFL, CDKN1A, ASS1, PER1 and PIK3R1
GO:1901698	response to nitrogen compound	$2.63x10^{-6}$	$1.87x10^{-5}$	GAL, DUSP1, NEFL, CDKN1A, ASS1, PER1 and PIK3R1
GO:0009416	response to light stimulus	$2.73x10^{-6}$	$1.85x10^{-5}$	DUSP1, CDKN1A, CRY2, PER1 and PIK3R1
GO:0033144	negative regulation of intracellular steroid hormone receptor signalling pathway	$3.18x10^{-6}$	$2.07x10^{-5}$	ARNTL, CRY2 and PER
GO:0051414	response to cortisol	$3.78x10^{-6}$	$2.36x10^{-5}$	SLIT3 and KLF9
GO:0048583	regulation of response to stimulus	$5.61x10^{-6}$	$3.36x10^{-5}$	ARNTL, DKK3, C1QTNF1, GAL, SLIT3, DUSP1, CRY2, CDKN1A, STAT5B, PER1, PIK3R and, IL6R
GO:0050896	response to stimulus	$9.46x10^{-6}$	$5.46x10^{-5}$	ARNTL, SLIT3, NEFL, CDKN1A, CRY2, ASS1, STAT5B, PER1, PIK3R1, GAL, DUSP1, KLF9 and IL6R

3.3.2 PRS prediction of MDD risk

The amount of variance in MDD explained by each effect in comparison to genome-wide prediction in regard to sample size across genesets (i.e. subsamples of genome-wide dataset with different number of SNPs) is shown in **Figure 3.2**. The pattern shows a higher risk of MDD explained within genesets than that expected in regard to their sample size (i.e. the number of SNPs). Particularly, for stress-sensitivity in geneset 3, the set of genes overlapping a glucocorticoid receptor binding sequence derived from a neuronal cell line. Given the number of SNPs that each geneset represent from the entire genome-wide dataset (**Table 3.1**) and the variance explained by this genome-wide dataset (reported in **chapter 2**), we estimated the variance expected to be explained by each dataset if such risk of MDD was homogeneously distributed along the genome. The enrichment of risk of MDD within each glucocorticoid-related geneset was estimated as the ratio between the risk observed and the risk expected (see **Table 3.6** and **Table 3.7**). Therefore, if the risk was not clustered in specific loci we would expect an enrichment of x1. No significant enrichment was found in any geneset using the stress-sensitivity PRS (PRS_{SS}) alone. (GS G1: x9.52, Empirical-*p* = 0.797; GS G2: x1.42, Empirical-*p* = 0.265; GS G3: x6.00, Empirical-*p* = 0.575; UKB G1: x15.64, Empirical-*p* = 0.880; UKB G2: x1.20, Empirical-*p* = 0.796; UKB G3: x18.41, Empirical-*p* = 0.274; **Table 3.6** and **Table 3.7**). The genetic contribution of PRS_D weighted by additive main effects of MDD or PRS_N weighted by additive main effects of neuroticism on MDD risk was only significant in geneset 2 (the largest set; ~40% of the genome-wide data), predicting on UKB (GS as discovery sample; PRS_D: x1.44, Empirical-*p* = 0.004; PRS_N: x2.02, Empirical-*p* = 0.002).

Enrichment of the full MDD risk conferred by the combined contributions of risk conferred by PRS_{MDD}, PRS_N and PRS_{SS} effects together was significant in both cohorts only on “down-stream” cortisol release sets (genesets 2 and 3), but not in the “up-stream” geneset (geneset 1), suggesting that MDD risk is found and enriched in glucocorticoid response genes but not in glucocorticoid signalling genes (see **Table 3.7**).

Enrichment of MDD risk explained by $PRS_{SS,}$ above that already explained by PRS_D and PRS_N combined together (model 5 vs. model 4) was significant on "down-stream" glucocorticoid response genes from genesets 2 and 3 in one cohort but not on the other. Enrichment was significant for geneset 2 in GS (x1.4, Empirical-*p* = 0.027) but this did not cross-validate in UKB (x1.26, Empirical-*p* = 0.147). Conversely, geneset 3 was significantly enriched in UKB (x19.6, Empirical-*p* = 0.034) but it did not cross-validate in GS (x6.18, Empirical-*p* = 0.056). As with the combined MDD risk conferred by all PRS effects together, these findings suggest that stress-sensitivity effect is enriched in glucocorticoid response genes, particularly in those found in neuronal cells (geneset 3). Such enrichment is detected from the remaining variance after taking into account the additive main contributions of MDD and neuroticism (see **Table 3.7**). The empirical cumulative distributions derived to estimate the empirical *p*-values are shown in **Figure 3.3**.

Figure 3.2 Risk of MDD explained by sample size. *Distribution of risk of MDD explained by sample size in comparison to the genome-wide predictions. Black dashed line represents an homogeneous distribution where the amount of variance in MDD explained is proportional to the sample size. Lines represent contributions from each PRS effect: MDD (blue; model 1 – null model), neuroticism (green; model 2 – null model) and stress-sensitivity (red; model 3 – null model). Orange line shows the risk of MDD explained by all three effects combined (model 5 – null model), and dashed red line is stress-sensitivity from the residual variation explained after taking into account the effects of MDD and neuroticism (model 5 – model 4). For each effect, data points compare the amount of MDD risk explained in regard to the amount of MDD risk explained under the same model by the same component in the genome-wide set (lines are independent from each other). G1 stands for geneset 1 (genes ontologically related with HPA axis; ~2.3% genome-wide sample), G2 stands for geneset 2 (glucocorticoid-response genes derived from human lung epithelial carcinoma; ~39.5% genome-wide sample), and G3 stands for geneset 3 (glucocorticoid-response genes derived from PC12 cells; ~5.7% genome-wide sample). Y-axis: the % of MDD risk explained in comparison to genome-wide prediction. X-axis: the number of SNPs in each set (%) in comparison to genome-wide dataset.*

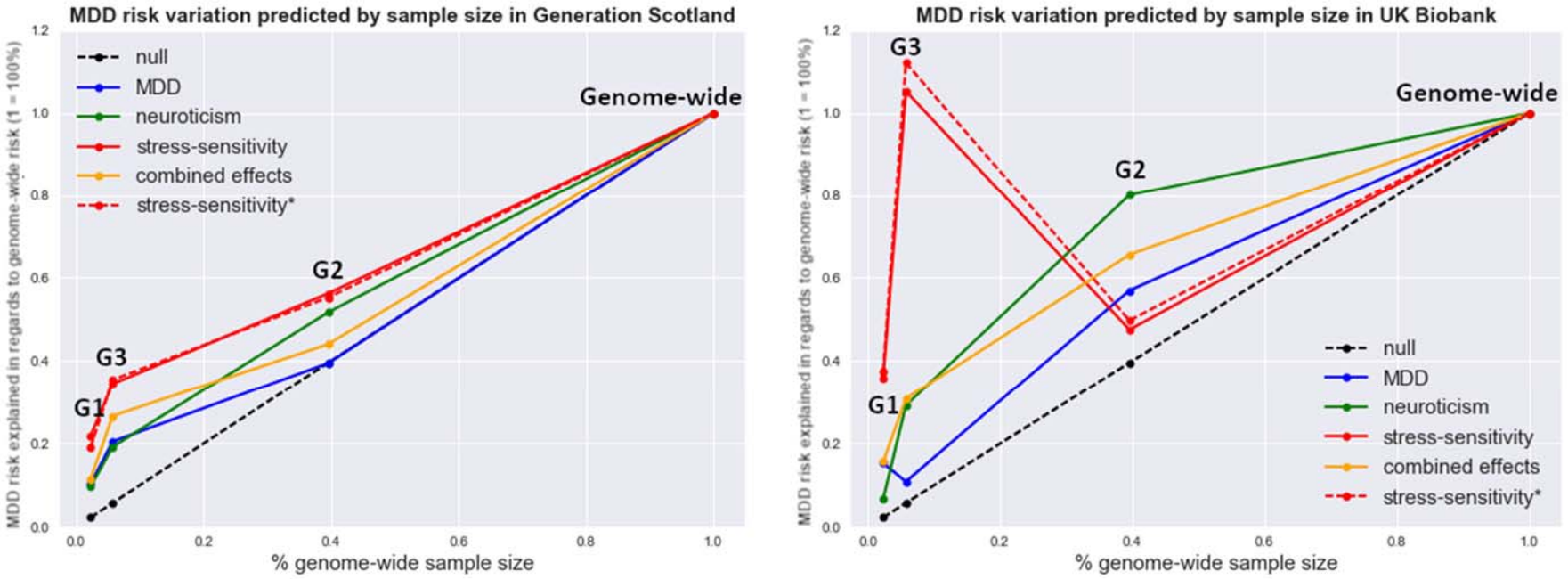


Table 3.6 Enrichment of genetic contributions to risk of depression within glucocorticoid-related genesets

Single models.

Target sample	Generation Scotland								
Genetic effect	**MDD**			**Neuroticism**			**Stress-sensitivity**		
Geneset	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**
Enrichment	x4.49	x1.00	x3.57	x4.20	x1.31	x3.35	x9.52	x1.42	x6.01
Observed *p*-value	0.206	0.013	0.075	0.251	7.91×10^{-3}	0.106	0.144	0.019	0.067
Empirical *p*-value	0.914	0.209	0.797	0.905	0.119	0.554	0.797	0.265	0.575
Target sample	**UK Biobank**								
Genetic effect	**MDD**			**Neuroticism**			**Stress-sensitivity**		
Geneset	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**
Enrichment	x6.64	**x1.44**	x1.90	x2.93	**x2.02**	x5.10	x15.64	x1.20	x18.41
Observed *p*-value	0.057	**2.35×10^{-4}**	0.109	0.245	**6.02×10^{-5}**	0.015	0.219	0.157	0.035
Empirical *p*-value	0.382	**4×10^{-3}**	0.546	0.821	**2×10^{-3}**	0.131	0.880	0.796	0.274

G 1 (geneset 1): genes ontologically related with HPA axis. G 2 (geneset 2): genes overlapping a GR binding sequence or within 50Kb from an A549 cell line (human lung epithelial carcinoma). G 3 (geneset 3): genes overlapping a GR binding sequence or within 50Kb from a PC12 cell line (pheochromoytoma of the rat adrenal medulla). Enrichment represents the ration between the MDD risk explained by each effect and that expected genome-wide MDD risk was homogeneously distributed along the genome. Observed *p*-value from a Likelihood ratio test assessing the significance of each PRS (PRS_{MDD}, PRS_{N} and PRS_{SS}; models 1, 2 and 3 against null model). Empirical *p*-value derived from an empirical cumulative distribution after permuting 10,000 times. In red: significant Empirical p-value < 0.05. In orange: 0.05 < Empirical p-value < 0.1.

Table 3.7 Enrichment of genetic contributions to risk of depression within glucocorticoid-related genesets

Combined models.

Target sample	Generation Scotland					
Genetic effect	**MDD, neuroticism and stress-sensitivity together**			**Stress-sensitivity (on top of MDD and neuroticism)**		
Geneset	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**
Enrichment	x5.03	**x1.11**	**x4.65**	x8.33	**x1.40**	x6.18
Observed *p*-value	0.230	**9.34×10^{-4}**	**0.019**	0.171	**0.020**	0.062
Empirical *p*-value	0.247	**0.001**	**0.016**	0.189	**0.027**	0.056
Target sample	**UK Biobank**					
Genetic effect	**MDD, neuroticism and stress-sensitivity together**			**Stress-sensitivity (on top of MDD and neuroticism)**		
Geneset	**G 1**	**G 2**	**G 3**	**G 1**	**G 2**	**G 3**
Enrichment	x6.79	**x1.66**	**x5.39**	x16.32	x1.26	**x19.60**
Observed *p*-value	0.098	**7.44×10^{-6}**	**5.99×10^{-3}**	0.217	0.155	**0.033**
Empirical *p*-value	0.089	**$\leq 1 \times 10^{-4}$**	**2×10^{-3}**	0.199	0.147	**0.034**

G 1 (geneset 1): genes ontologically related with HPA axis. G 2 (geneset 2): genes overlapping a GR binding sequence or within 50Kb from an A549 cell line (human lung epithelial carcinoma). G 3 (geneset 3): genes overlapping a GR binding sequence or within 50Kb from a PC12 cell line (pheochromoytoma of the rat adrenal medulla). Enrichment represents the ration between the MDD risk explained by each effect and that expected genome-wide MDD risk was homogeneously distributed along the genome. Observed *p*-value from a Likelihood ratio test assessing: (left side) the significance of all PRS together (PRS_{MDD}, PRS_{N} and PRS_{SS}; model 5 against model), and the significance of PRS_{SS} to explain MDD risk above that explained by both PRS_{MDD} and PRS_{N} together (model 5 against model 4). Empirical *p*-value derived from an empirical cumulative distribution after permuting 10,000 times. In red: significant Empirical p-value < 0.05. In orange: 0.05 < Empirical p-value < 0.1.

Figure 3.3a

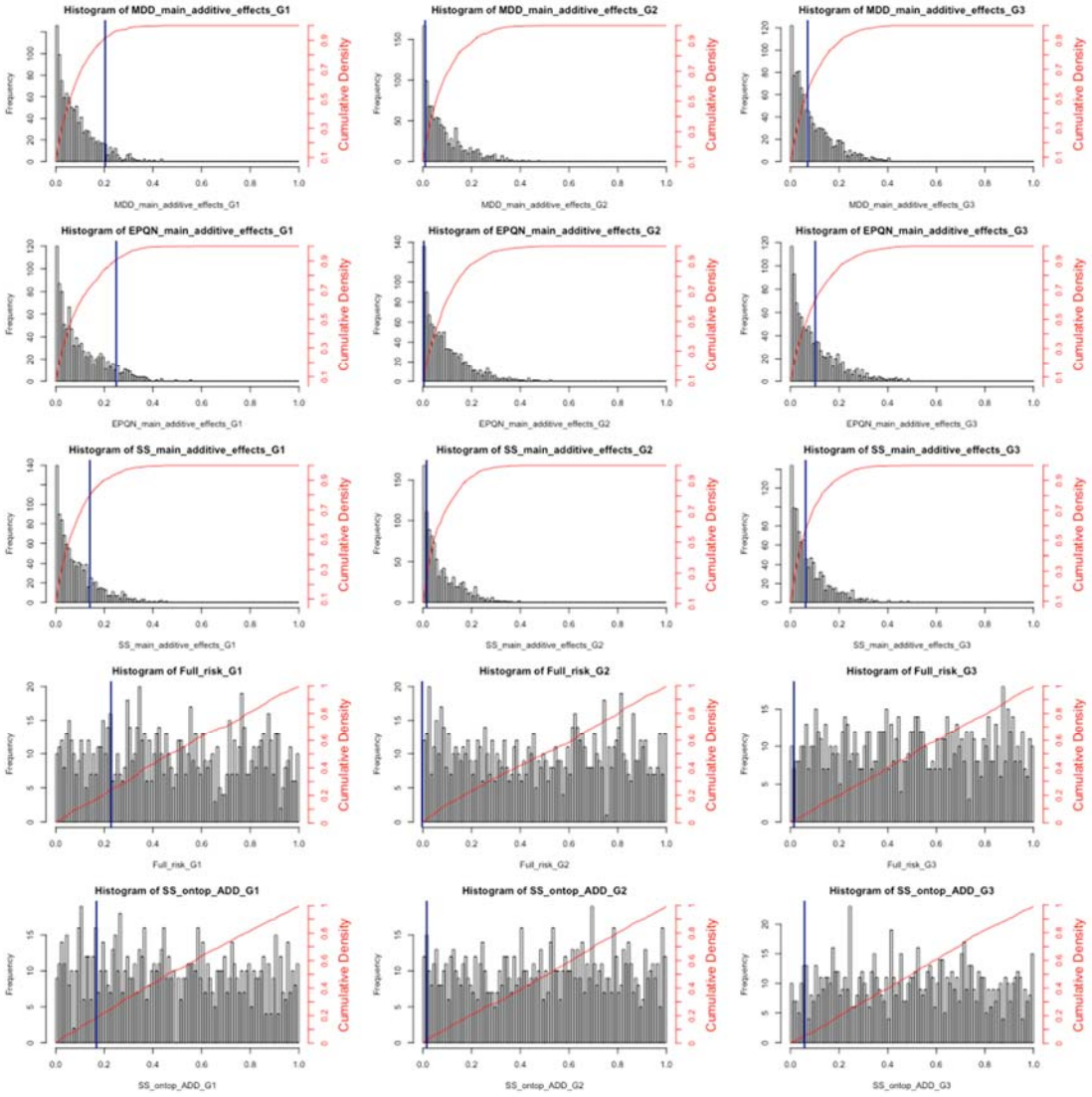

Histogram of MDD_main_additive_effects_G1
Histogram of MDD_main_additive_effects_G2
Histogram of MDD_main_additive_effects_G3
Histogram of EPQN_main_additive_effects_G1
Histogram of EPQN_main_additive_effects_G2
Histogram of EPQN_main_additive_effects_G3
Histogram of SS_main_additive_effects_G1
Histogram of SS_main_additive_effects_G2
Histogram of SS_main_additive_effects_G3
Histogram of Full_risk_G1
Histogram of Full_risk_G2
Histogram of Full_risk_G3
Histogram of SS_ontop_ADD_G1
Histogram of SS_ontop_ADD_G2
Histogram of SS_ontop_ADD_G3
Frequency
Cumulative Density

Figure 3.3b

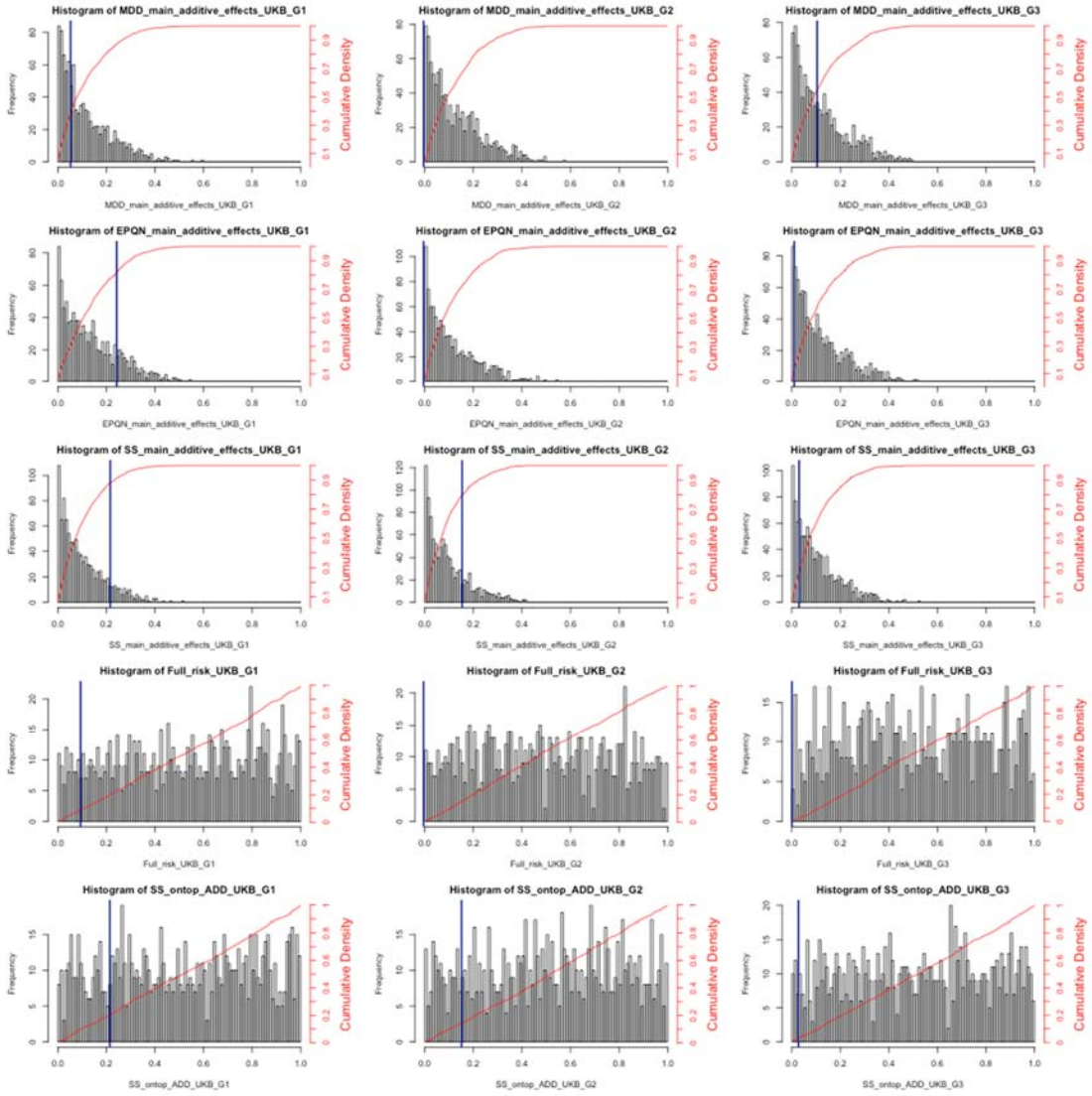

Histogram of MDD_main_additive_effects_UKB_G1
Histogram of MDD_main_additive_effects_UKB_G2
Histogram of MDD_main_additive_effects_UKB_G3
Histogram of EPQN_main_additive_effects_UKB_G1
Histogram of EPQN_main_additive_effects_UKB_G2
Histogram of EPQN_main_additive_effects_UKB_G3
Histogram of SS_main_additive_effects_UKB_G1
Histogram of SS_main_additive_effects_UKB_G2
Histogram of SS_main_additive_effects_UKB_G3
Histogram of Full_risk_UKB_G1
Histogram of Full_risk_UKB_G2
Histogram of Full_risk_UKB_G3
Histogram of SS_ontop_ADD_UKB_G1
Histogram of SS_ontop_ADD_UKB_G2
Histogram of SS_ontop_ADD_UKB_G3
Frequency
Cumulative Density

Figure 3.3. Empirical cumulative distributions in a) Generation Scotland and b) UK Biobank. *Results from geneset 1 are shown in first column, geneset 2 in second column, and geneset 3 in third column. The first three rows show an empirical cumulative distribution of p-values after permuting 1,000 times MDD from model 1 (1st row, PRS_{MDD} effect), model 2 (2nd row, PRS_{N} effect), and model 3 (3rd row, PRS_{SS} effect). The 4th row shows the empirical cumulative distribution of p-values derived to compare model 5 (which includes all PRS fitted together) against a null model. And the 5th row shows the empirical cumulative distribution of p-values derived to compare model 5 (which includes all PRS fitted together) against model 4 (which includes the main effects of MDD and neuroticism) to assess the significance of PRS_{SS} weighted by stress-sensitivity effect. All models were permuted 1,000 times. X-axis shows the p-value estimated by a likelihood ratio test. A vertical dark blue line indicates the observed p-values. Significance threshold for the empirical p-values was fixed at $\alpha = 0.05$.*

3.4 Discussion

In this chapter, I assessed whether genetic contributions to risk of MDD are enriched within set of genes related to the HPA axis and the glucocorticoid signalling and response pathway. Overall, results indicate that genetic contributions to risk of MDD, rather than being homogeneously distributed along the genome, are enriched within “down-stream” glucocorticoid response genes, but not within “up-stream” glucocorticoid signalling genes. Noteworthy, the genetic contribution to the stress-sensitivity trait derived in **chapter 2** was enriched in sets of glucocorticoid response genes when the additive main contributions to MDD and neuroticism were taken into account.

Three different glucocorticoid-related genesets were defined from HPA-related gene ontology terms (geneset 1) and glucocorticoid response expression studies on: human non-neuronal cell line (geneset 2)[367] and rat neuronal cell line (geneset 3)[368]. As expected, geneset 1 was enriched in genes responding to glucocorticoid pathways, mostly, with receptor binding functions in either extracellular or cytoplasmic parts, including neuron part. Genes overlapping glucocorticoid receptor binding sites in the neuronal cell line were enriched in biological process such as regulation of phosphate metabolism, single-organism process, developmental process, cellular process, cellular component movement and response to chemical; molecular functions such as protein kinase activity, voltage-gated cation channel activity, enzyme binding and anion binding; and in cell components such as synapse part, neuron part, cytoplasmic part and cell junction; showing the relation with the neuronal context where these genes come from. Notably, previous pathway analyses have also associated, among others, neuron and synapse parts, cell junctions or voltage-gated cation channels with depression[150,151,160]. Some ontologies such as positive regulation of biological process, anion binding function, cytoplasmic part and cell junction were similarly enriched in genes overlapping glucocorticoid receptor binding sites in both neuronal and non-neuronal cells, the latest with most enriched

biological process related to regulation of intracellular signal transduction. There was poor overlap between genes from each designed glucocorticoid-related geneset. This supports glucocorticoid receptor occupancy and, therefore, glucocorticoid receptor regulation being highly cell type specific, as it was already suggested by Polman *et al.*[368]. Gene-expression profiles induced by glucocorticoid receptors are cell-type specific, so genomic binding sites binding glucocorticoid receptors may differ from a cell-line to another. However, some degree of overlap was expected given the negative feedback loop between glucocorticoids and the HPA axis. At certain levels of cortisol released in response to stressful stimulus, cortisol exerts a negative feedback through glucocorticoid receptor signalling inhibiting the release of CRF in the hypothalamus and ACTH in the pituitary[362]. This regulatory mechanism restores hormone levels to normality when stress has ceased and maintains secretion within physiological levels, thus, protecting against the prolonged effects of stress. Noteworthy, the top biological process detected in the functional analysis of genes overlapping all glucocorticoid-related genesets was ”rhythmic process” with 7 of the 16 overlapping genes included. 5 of them (*ARNTL*, *CRY2*, *ASS1*, *PER1* and *KLF9*) were related to “circadian rhythm”. Stress has great influence in the circadian system[386], whose disruption shows robust effects on mood instability, adverse wellbeing outcomes and reduced cognitive function, including an increased risk of MDD[387]. Therefore, the circadian clock may mediate the association between stress and stress-related conditions.

We saw that the MDD risk predicted by glucocorticoid-related genesets deviated from that expected given the number of SNPs they included (**Figure 3.2**). The small difference between the stress-sensitivity effect alone and when combined with main additive effects of MDD and neuroticism suggests that genetic contributions to stress-sensitivity within glucocorticoid-related genes is independent of main contributions to MDD and neuroticism. The most substantial deviation and the greatest enrichment of stress-sensitivity effect (**table 3.6** and **table 3.7**) were found within geneset 3, which includes human orthologous genes overlapping or within 50 kb from a genomic

binding site derived by ChipSeq from a neuronal cell line from rat (i.e. PC12 cells). This geneset, that only included 5.7% of the genome-wide SNPs, showed a significant prediction of risk of MDD more than 19 times higher than expected in UKB. This enrichment also approached significance in GS (x6.18, Empirical-*p* = 0.056). Although this neuronal cell line come from rat and not human, finding highlights the need of targeting approaches like this in neuronal cell types. Given the effects of glucocorticoids on neuronal plasticity and brain functioning[350,351], the use of data on genomic binding sites binding glucocorticoid receptor, and on its glucocorticoid response elements, characterized from neuronal tissues coming from human should be an essential key point to study genetic stress responses determining glucocorticoid signalling underlying the risk of depression. However, Chip-seq data on glucocorticoid binding sites from human neuronal-lines was not available.

The results suggest that genes within “down-stream” glucocorticoid response genesets explain more variation in MDD risk than expected. The gene-based test analysis from **chapter 2** identified *ZNF366,* which may, directly or indirectly, effect glucocorticoid receptor-regulated gene expression, suggesting that the genetic component identified affecting sensitivity to environmental stress (so-called stress-sensitivity) may act, although likely not exclusively, through glucocorticoid receptor-related pathways and extensively, the HPA axis and cortisol signalling. However, the results presented in this chapter suggest that such stress-sensitivity effect is found in “down-stream” glucocorticoid-response genes but not in “up-stream” signalling genes involved in the HPA axis. Several polymorphisms from genes associated to the HPA axis such as *CRH1*, *FKBP5, NR3C1* and *NR3C2* have been investigated as potential modulators of the effect of stress on depression[372-375,388,389]. The interaction of early life stress and PRS constructed using the aggregated number of high-risk alleles from some of these polymorphisms have been report to predict cortisol levels, hippocampal volumes and amygdala function[376,390], and to increase depressive symptoms in offspring of depressed mothers, with the greatest effect in those

experiencing the highest levels of stress[375]. Thus, suggesting that genetic variants from genes related to the HPA axis may moderate the association between stress reactivity and depression. *CRH1* and *NR3C2* were included in genes ontologically-related to the HPA axis in geneset 1, and *FKBP5* overlapped with glucocorticoid receptor binding sites in the human non-neuronal cell line and thus, included in geneset 2. *NR3C1,* as well as *BDNF,* another candidate gene widely investigated in previous GxE studies[391-393], were also included in geneset 1 and geneset 2. However, none of these genes overlapped glucocorticoid receptor binding sites tested in a rat neuronal cell line (i.e. not included in geneset 3). Findings suggest to target glucocorticoid response genes rather than genes involved in the HPA axis as candidate genes for GxE studies. An enrichment of stress-sensitivity effect (i.e. genetic contributions to the MDD-depended change in neuroticism) contributing to risk of MDD was seen in the "down-stream" sets of glucocorticoid response genes from the residual variance after taking into account the additive main effects of both MDD and neuroticism (the genetic contributions to the stable component of neuroticism), supporting gene-environment interaction effects in the glucocorticoid response pathway. Such enrichment of MDD risk and stress-sensitivity effect in glucocorticoid response genes may help to explain why more than 70% of patients with active Cushing's syndrome, which is caused by excessive endogenous levels of circulating free cortisol, report co-morbid disorders ranging from anxiety to psychosis (stress-related conditions), with MDD being the most common[394]; or patients with Adison's disease, which is characterized by low cortisol levels, report depression as a significant co-morbidity[395].

In a previous study, Arloth *et al.* suggested that the risk of developing MDD after adverse life events may be influenced by an individual's sensitivity to the transcriptional effects of cortisol released during the stressful adverse events[377]. Using a stimulated eQTL approach, they highlighted that one putative biological mechanism underlying the interaction of both genetic and environmental risk factors conferring risk of depression may implicates variability in the glucocorticoid regulation induced by cortisol in response of

stress. My results support such findings from Arloth *et al.* and suggest that, at least in a subset of patients, such individual's sensitivity to stress could depend on a specific genetic architecture distinct from the additive main contributions to an individual's risk detected in GWAS of either MDD or neuroticism.

There are few caveats to consider in the analysis presented in this chapter. First, many different approaches to define glucocorticoid-related genesets could be applied, providing sets of different sizes and thus being more or less representative of the entire genome. For example, more "up-stream" glucocorticoid signalling set of genes that may reflect better HPA axis activity than those genes ontologically related to the HPA axis could be defined using genes expected to explain cortisol secretion[396] or genes previously linked to the regulation of the HPA axis activity identified in other studies[397]. In addition, although many glucocorticoid receptor likely bind glucocorticoid response elements within promoter or enhancer regions of a gene and thus stimulating its transcription, other genes modulated by glucocorticoid receptor may be far from such genomic binding sites. Therefore, taking the overlapping or closest gene to each genomic binding site may not provide the best candidate gene to be included in all regions. Furthermore, using genotype data on hundreds of thousands of SNPs, rather than imputed or whole-genome sequencing data on millions of SNPs, followed by a round of clumping to select independent SNPs in linkage equilibrium resulted in a very low number of SNPs representative of each gene. Some loci with a glucocorticoid receptor binding site may even lack of SNP representation, as seen in few genes without SNP data in GS, resulting in a loss of information. For example, there was a lack of SNPs covering *MAOA* (**Appendix B**: Supplementary Table 2), a gene overlapped with glucocorticoid receptor binding sites in the human cell line and a candidate to interact with stress in previous GxE studies[398,399]. A less stringent clumping step would reduce the loss of informative data in detrimental of linkage equilibrium across SNPs.

In summary, we detected more genetic variation in risk of MDD within sets of glucocorticoid response genes than expected, and particularly in MDD risk conferred by the MDD-dependent stress-sensitivity effect derived in **chapter 2**, independently of main additive contributions of MDD. The results presented suggest that the association between environmental stress and MDD may be mediated by an individual's sensitivity to the transcriptional effects of glucocorticoid receptors following cortisol release during stressful events, as suggested by Arloth *et al.*[377]. A deeper understanding on the relationship between psychological stress and MDD through adaptive changes in stress response would facilitate the discovery of novel pathways and thus identify new targets for drug discovery.

Chapter 4 A validation of the diathesis-stress model for depression in Generation Scotland

The cornerstone of the preceding chapters was a proxy for sensitivity to stress derived without the requirement of direct measures of SLE. In contrast, in order to study GxE effects underlying liability to major depressive disorder, in the following chapters, I incorporate variables to measure the number of SLE occurring during a specific time period prior to assessment of depressive symptoms. In this chapter and the following ones, I use quantitative scores to assess depression, rather than MDD diagnosis.

As detailed in the *Background* **chapter 1** (section 1.6.4.3), Colodro-Conde *et al.* reported a significant PRSxSLE effect on liability to depression under a *diathesis-stress* model. In this chapter, I validate this result by testing the *diathesis-stress* model for depression in a subsample of Generation Scotland that took part in a longitudinal follow-up study including assessment of SLE. To properly replicate Colodro-Conde *et al.* study, the same instruments to assess GxE effects should be applied, but such instruments are not available in Generation Scotland. However, I apply the same methodology and *diathesis-stress* framework as the original study, adapting my study to the available instruments that attempt to capture the same underlying context.

Measures of recent SLE were collected through a brief life events questionnaire based on the List of Threatening Experiences (the questionnaire can be seen in **Appendix D.5**). The data reported was used to construct variables of SLE from the 6 months preceding a self-reported questionnaire of depressive symptoms. This data was collected using the General Health Questionnaire (GHQ), a reliable and validated psychometric screening tool developed by Goldberg and Hillier in 1972, which takes less than 5 minutes to complete, and consists of four subscales (somatic symptoms, anxiety and insomnia, social dysfunction, and severe depression)

designed to detect common psychiatric disorders, or current mental distress, on adults and adolescents (but not children) by non-psychiatric clinical professionals (e.g. researchers). The General Health Questionnaire is copyrighted, therefore, I cannot show the questions involved. The use of the questionnaire is licensed by GL Assessment (https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/). A license agreement must be completed beforehand and a user fee is required to all users (commercial and academic users).

This chapter has been published in *Translational Psychiatry* and is presented as submitted, which explains the use of "we" within the chapter. I confirm that the work of this chapter is my own work under guidance from my supervisor Dr. Pippa Thomson. I performed all the analyses myself. The published article and **Supplementary Material** can be found in **Appendix C**.

Publication:

Arnau-Soler, A. *et al.* A validation of the diathesis-stress model for depression in Generation Scotland. *Translational Psychiatry* 9, 25 (2019).

4.1 Abstract

Depression has well-established influences from genetic and environmental risk factors. This has led to the *diathesis-stress* theory, which assumes a multiplicative gene-by-environment interaction (GxE) effect on risk. Recently, *Colodro-Conde et al.* empirically tested this theory, using the polygenic risk score for major depressive disorder (PRS, genes) and stressful life events (SLE, environment) effects on depressive symptoms, identifying significant GxE effects with an additive contribution to liability.

We have tested the *diathesis-stress* theory on an independent sample of 4,919 individuals.

We identified nominally significant positive GxE effects in the full cohort (R^2 = 0.08%, p = 0.049) and in women (R^2 = 0.19%, p = 0.017), but not in men (R^2 = 0.15%, p = 0.07). GxE effects were nominally significant, but only in women, when SLE were split into those in which the respondent plays an active or passive role (R^2 = 0.15%, p = 0.038; R^2 = 0.16%, p = 0.033, respectively). High PRS increased the risk of depression in participants reporting high numbers of SLE (p = 2.86 x 10^{-4}). However, in those participants who reported no recent SLE, a higher PRS appeared to increase the risk of depressive symptoms in men (β = 0.082, p = 0.016) but had a protective effect in women (β = -0.061, p = 0.037). This difference was nominally significant (p = 0.017).

Our study reinforces the evidence of additional risk in the aetiology of depression due to GxE effects. However, larger sample sizes are required to robustly validate these findings.

4.2 Introduction

Stressful life events (SLE) have been consistently recognized as a determinant of depressive symptoms, with many studies reporting significant associations between SLE and major depressive disorder (MDD)[84,90,93-95,400,401]. Some studies suggest that severe adversity is present before the onset of illness in over 50% of individuals with depression[402] and may characterize a subtype of cases[403]. However, some individuals facing severe stress never present symptoms of depression[404]. This has led to a suggestion that the interaction between stress and an individual's vulnerability, or *diathesis*, is a key element in the development of depressive symptoms. Such vulnerability can be conceived as a set of biological factors that predispose to illness. Several *diathesis-stress* models have been successfully applied across many psychopathologies[225,405-408].

The *diathesis-stress* model proposes that a latent *diathesis* may be activated by stress before psychopathological symptoms manifest. Some levels of *diathesis* to illness are present in everybody, with a threshold over which symptoms will appear. Exceeding such a threshold depends on the interaction between *diathesis* and the degree of adversity faced in SLE, which increases the liability to depression beyond the combined additive effects of the *diathesis* and stress alone[225]. Genetic risk factors can, therefore, be conceived as a genetic *diathesis*. Thus, this genetically driven effect produced by the *diathesis-stress* interaction can be seen as a gene-by-environment interaction (GxE).

MDD is characterized by a highly polygenic architecture, composed of common variants with small effect and/or rare variants[150]. Therefore, interactions in depression are also expected to be highly polygenic. In recent years, with the increasing success of genome-wide association studies, GxE studies in depression have shifted towards hypothesis-free genome-wide and polygenic approaches that capture liability to depression using genetic

data[83,86,113,169,267-269,275,276]. Recent advances in genomics and the massive effort from national institutions to collect genetic, clinical and environmental data on large population-based samples now provide an opportunity to empirically test the *diathesis-stress* model for depression. The construction of polygenic risk scores (PRS) offers a novel paradigm to quantify genetic *diathesis* into a single genetic measure, allowing us to study GxE effects with more predictive power than any single variant[409-412]. PRS are genetic indicators of the aggregated number of risk alleles carried by an individual weighted by their allelic effect estimated from genome-wide association studies. This polygenic approach to assessing the *diathesis-stress* model for depression has been tested using either childhood trauma[83,86,275] or adult SLE[86,169,276] as measures of environmental adversity.

Recently, Colodro-Conde *et al.*[169] provided a direct test of the *diathesis-stress* model for recent SLE and depressive symptoms. In this study, Colodro-Conde *et al.* used PRS weighted by the most recent genome-wide meta-analysis conducted by the Psychiatric Genetics Consortium (PGC; N = 159,601), and measures of three environmental exposures: lack of social support, "personal" SLE, and "network" SLE. Colodro-Conde *et al.* reported a significant additive risk on liability to depression due to a GxE effect in individuals who combine a high genetic predisposition to MDD and a high number of reported "personal" SLE, mainly driven by effects in women. A significant effect of interaction was not detected in males. They found no significant interaction between the genetic *diathesis* and "network" SLE or social support. They concluded that the effect of stress on risk of depression was dependent on an individual's *diathesis,* thus supporting the *diathesis-stress* theory. In addition, they suggested possible sex-specific differences in the aetiology of depression. However, Colodro-Conde *et al.* findings have not, to our knowledge, been independently validated.

In the present study we aim to test the *diathesis-stress* model in an independent sample of 4,919 unrelated white British participants from a further longitudinal follow-up from Generation Scotland and assess the

differences between women and men, using self-reported depressive symptoms and recent SLE.

4.3 Materials and methods

4.3.1 Sample description

Generation Scotland is a family-based population cohort recruited throughout Scotland by a cross-disciplinary collaboration of Scottish medical schools and the National Health Service (NHS) between 2006 and 2011[305]. At baseline, blood and salivary DNA samples from Generation Scotland participants were collected, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. Genome-wide genotype data were generated using the Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry[413]. The procedures and further details for DNA extraction and genotyping have been extensively described elsewhere[307,414]. In 2014, 21,525 participants from Generation Scotland eligible for re-contact were sent self-reported questionnaires as part of a further longitudinal assessment funded by a Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL)[415] to collect new and updated mental health questionnaires including psychiatric symptoms and SLE measures. 9,618 re-contacted participants from Generation Scotland agreed to provide new measures to the mental health follow-up[415] (44.7% response rate). Duplicate samples, those showing sex discrepancies with phenotypic data, or that had more than 2% missing genotype data, were removed from the sample, as were samples identified as population outliers in principal component analysis (mainly non-Caucasians and Italian ancestry subgroup). In addition, individuals with diagnoses of bipolar disorder, or with missing SLE data, were excluded from the analyses. SNPs with more than 2% of genotypes missing, Hardy-Weinberg Equilibrium test $p < 1 \times 10^{-6}$, or a minor allele frequency lower than 1%, were excluded. Individuals were then filtered by degree of relatedness (pi-hat < 0.05) using PLINK v1.9[273], maximizing retention of those participants reporting higher numbers of SLE (see phenotype assessment below). After quality control, the final dataset comprised 4,919 unrelated

individuals of European ancestry and 560,351 SNPs (mean age at questionnaire: 57.2, s.d. = 12.2, range 22-95; *women*: *n* = 2,990 - 60.8%, mean age 56.1, s.d. = 12.4; *men*: *n* = 1,929 - 39.2%, mean age 58.7, s.d. = 11.8). Further details on the recruitment procedure and Generation Scotland profile are described in detail elsewhere[124,304-307]. All participants provided written consent. All components of Generation Scotland and STRADL obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the National Health Service (reference 05/s1401/89). Generation Scotland data is available to researchers on application to the Generation Scotland Access Committee (access@generationscotland.org).

4.3.2 Phenotype assessment

Participant self-reported current depressive symptoms through the 28-item scaled version of The General Health Questionnaire[416,417]. The General Health Questionnaire is a reliable and validated psychometric screening tool to detect common psychiatric and non-psychotic conditions (General Health Questionnaire Cronbach alpha coefficient: 0.82 – 0.86)[197]. This consists of 28 items designed to identify whether an individual's current mental state has changed over the last 2 weeks from their typical state. The questionnaire captures core symptoms of depression through subscales for severe depression, emotional (e.g. anxiety and social dysfunction) and somatic symptoms linked to depression. These subscales are highly correlated[418] and suggest an overall general factor of depression[419]. Participants rated the 28 items on a four-point Likert scale from 0 to 3 to assess its degree or severity[197] (e.g., *Have you recently felt that life is entirely hopeless?* "Not at all", "No more than usual", "Rather more than usual", "Much more than usual"), resulting on an 84-point scale depression score. The Likert scale, which provides a wider and smoother distribution[197], could be more sensitive to detect changes in mental status in those participants with chronic conditions or chronic stress who may feel their current symptoms as "usual"[420], and to detect psychopathology changes as response to stress. The final depression score was log transformed to reduce the effect of positive skew and provide a better approximation to a normal distribution. In

addition, participants completed the Composite International Diagnostic Interview–Short Form, which diagnoses lifetime history of MDD according to DSM-IV criteria[421]. The depression score predicted lifetime history of MDD (odd ratio = 1.91, 95% confidence intervals 1.80-2.02, $p = 1.55 \times 10^{-102}$, N = 8,994), with a 3.8-fold increased odd of having a lifetime history of MDD between participants in the top and bottom deciles, thus supporting the usefulness of the depression score in understanding MDD. For a better interpretation, we scaled the depression score to a mean of 0 when required (**Figure 4.3**).

Data from a self-reported questionnaire based on the List of Threatening Experiences[99] was used to construct a measure of common SLE over the previous 6 months. The List of Threatening Experiences is a reliable psychometric device to measure psychological "stress"[422,423]. It consists of a 12-item questionnaire to assess SLE with considerable long-term contextual effects (e.g., *Over last 6 months, did you have a serious problem with a close friend, neighbour or relatives?*). A final score reflecting the total number of SLE (TSLE) ranging from 0 to 12 was constructed by summing the "yes" responses. Additionally, TSLE was split into two categories based on those items measuring SLE in which the individual may play and active role exposure to SLE, and therefore in which the SLE is influenced by genetic factors and thus subject to be "dependent" on an individual's own behaviour or symptoms (DSLE; 6 items, e.g., *a serious problem with a close friend, neighbour or relatives* may be subject to a respondent's own behaviour), or SLE that are not influenced by genetic factors, likely to be *independent* on a participant's own behaviour (ISLE; 5 items, e.g., *a serious illness, injury or assault happening to a close relative* is potentially independent of a respondent's own behaviour)[99,100]. The item "*Did you/your wife or partner give birth?"* was excluded from this categorization. In addition, SLE reported were categorized to investigate the *diathesis* effect at different levels of exposure, including a group to test the *diathesis* effect when SLE is not reported. 3 levels of SLE reported were defined (0 SLE = "none", 1 or 2 SLE

= “low”, and 3 or more SLE = “high") to retain a large enough sample size for each group to allow meaningful statistical comparison.

4.3.3 Polygenic profiling & statistical analysis

Polygenic risk scores (PRS) were generated by PRSice[313], whose functionality relies mostly on PLINK v1.9[273], and were calculated using the genotype data of Generation Scotland participants (i.e. target sample) and summary statistics for MDD from the PGC-MDD2 GWAS release (July 2016, discovery sample) used by Colodro-Conde *et al*.[169], with the added contribution from QIMR cohort and the exclusion of Generation Scotland participants, resulting in summary statistics for MDD derived from a sample of 50,455 cases and 105,411 controls.

Briefly, PRSice removed strand-ambiguous SNPs and clump-based pruned (r^2 = 0.1, within a 10Mb window) our target sample to obtain the most significant independent SNPs in approximate linkage equilibrium. Independent risk alleles were then weighted by the allelic effect sizes estimated in the independent discovery sample and aggregated into PRS. PRS were generated for eight *p* thresholds (*p* thresholds: < 5 x 10^{-8}, < 1 x 10^{-5}, < 0.001, < 0.01, < 0.05, < 0.1, < 0.5, <=1) determined by the discovery sample and standardized (See **Appendix C**: Supplementary Table 1 for summary of PRS).

A genetic relationship matrix (GRM) was calculated for each dataset (i.e. *full cohort*, *women* and *men*) using GCTA 1.26.0[176]. Mixed linear models using the GRM were used to estimate the variance in depression score explained by PRS, SLEs and their interaction; and stratified by sex. 20 principal components were calculated for the datasets.

The mixed linear model used to assess the effects of PRS is as follows:

$$Depression = \beta_0 + \beta_1 PRS + GRM + Covariates$$

Mixed linear models used to assess the effect of the stressors are as follows:

$$Depression = \beta_0 + \beta_1 TSLE + GRM + Covariates$$

$$Depression = \beta_0 + \beta_1 DSLE + GRM + Covariates$$

$$Depression = \beta_0 + \beta_1 ISLE + GRM + Covariates$$

Following Colodro-Conde *et al.*[169], covariates (i.e. age, age^2, sex, age-by-sex and age^2-by-sex interactions, and 20 principal components) were regressed from PRS (PRS') and SLE scores (i.e. TSLE', DSLE' and ISLE'; SLEs') before fitting models in GCTA to guard against confounding influences on the PRS-by-SLEs interactions[424]. PRS' and SLEs' were standardized to a mean of 0 and a standard deviation of 1. The Mixed linear models (i.e. the *diathesis-stress* model) used to assess GxE effects are as follows:

$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 TSLE' + \beta_3 PRS' x TSLE' + GRM + Covariates$$

$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 DSLE' + \beta_3 PRS' x DSLE' + GRM + Covariates$$

$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 ISLE' + \beta_3 PRS' x ISLE' + GRM + Covariates$$

Covariates fitted in the models above were age, age^2, sex, age-by-sex, age^2-by-sex and 20 principal components. Sex and its interactions (age-by-sex and age^2-by-sex) were omitted from the covariates when stratifying by sex. All parameters from the models were estimated using GCTA and the significance of the effect (β) from fixed effects assessed using a Wald test. The significance of main effects (PRS and SLEs) allowed for nominally testing the significance of interactions at *p*-threshold = 0.05. To account for multiple testing correction, a Bonferroni's adjustment correcting for 8 PRS and 3 measures of SLE tested (24 tests) was used to establish a robust threshold for significance at $p = 2.08 \times 10^{-3}$.

The PRS effect on depression score at different levels of exposure was further examined for the detected nominally significant interactions by categorizing participants on three groups based on the number of SLE reported (i.e. "none", "low" or "high"). Using linear regression, we applied a least squares approach to assess PRS' effects on the depression score in

each SLE category. Further conservative Bonferroni correction to adjust for the 3 SLE categories tested established a threshold for significance of p = 6.94×10^{-4}.

Differences on the estimated size of GxE effect between women and men were assessed by comparing a z-score to the standard normal distribution (α = 0.05, one-tailed). Z-scores were derived from GxE estimates (β) and standard errors (SE) detected in women and men as follows:

$$\mathrm{Z-score} = \frac{\beta_{women} - \beta_{men}}{\sqrt{SE(\beta_{women})^2 + SE(\beta_{men})^2}}$$

4.4 Results

PRS for MDD significantly predicted the depression score across the whole sample (β = 0.080, s.e. = 0.014, p = 7.53 x 10^{-9}) explaining 0.64% of the variance at its best p-threshold (p-threshold = 0.1; **Figure 4.1a**). Stratifying by sex, PRS significantly predicted the depression score in both sexes, explaining 0.59% in men and 0.67% in women (*men*: p-threshold = 0.1, β = 0.077, s.e. = 0.022, p = 2.09 x 10^{-4}; *women*: p-threshold = 0.1, β = 0.082, s.e. = 0.018, p = 4.93 x 10^{-6}; **Figure 4.1a**). Self-reported SLE over the last 6 months (TSLE, mean = 1.3 SLE, s.d. = 1.5) also significantly predicted depression score for the whole sample and stratified by sex (*full cohort*: variance explained = 4.91%, β = 0.222, s.e. = 0.014, p = 9.98 x 10^{-59}; *men*: 4.19%, β = 0.205, s.e. = 0.021, p = 2.23 x 10^{-22}; *women*: 5.33%, β = 0.231, s.e. = 0.018, p = 7.48 x 10^{-38}; **Figure 4.1b**). Overall, significant additive contributions from genetics and SLE in depression score were detected in all participants and across sexes. There was no significant difference in the direct effect of TSLE between women and men (p = 0.17). However, the variance in depression score explained by the TSLE appeared to be lower than the variance explained by the measure of personal SLE (PSLE) used in Colodro-Conde *et al.*[169] (12.9%). This may, in part, be explained by different contributions of dependent and independent SLE items screened in Colodro-Conde *et al.* compared to our study. Although questions about dependent SLE (DSLE, mean = 0.4 SLE) represented over 28% of the TSLE-items reported in our study, the main effect of DSLE explained approximately 93% of the amount of variance explained by TSLE (*full cohort*: variance explained = 4.56%, β = 0.212, s.e. = 0.014, p = 1.73 x 10^{-54}; *men*: 3.74%, β = 0.193, s.e. = 0.021, p = 9.66 x 10^{-21}; *women*: 5.07%, β = 0.225, s.e. = 0.018, p = 8.09 x 10^{-35}; **Figure 4.1b**). Independent SLE (ISLE, mean = 0.85 SLE), which represented over 69% of TSLE-items, explained approximately 57% of the amount of variance explained by TSLE (*full cohort*: variance explained = 2.80%, β = 0.167, s.e. = 0.014, p = 1.32 x 10^{-33}; *men*: 2.44%, β = 0.156, s.e. = 0.022, p = 2.88 x 10^{-13}; *women*: 3.02%, β = 0.174, s.e. = 0.018, p = 5.20 x

10^{-22}; **Figure 4.1b**). To explore the contribution from each measure, we combined DSLE and ISLE together in a single model. DSLE explained 3.34% of the variance of depressive score compared to 1.45% of the variance being explained by ISLE, suggesting that DSLE have a greater effect on liability to depressive symptoms than ISLE.

A *diathesis-stress* model for depression was tested to assess GxE effects. We detected significant, albeit weak, GxE effects on depression score (**Figure 4.2**). The PRS interaction with TSLE was nominally significant in the full cohort (β = 0.028, s.e. = 0.014, R^2 = 0.08%, *p* = 0.049) and slightly stronger in women (β = 0.044, s.e. = 0.018, R^2 = 0.19%, *p* = 0.017; **Figure 4.2a**), compared to men in which the effect was not significant (β = 0.039, s.e. = 0.022, R^2 = 0.15%, *p* = 0.07). However, these results did not survive correction for multiple testing ($p > 2.08 \times 10^{-3}$).

The best-fit threshold was much lower in women (*p*-threshold = 1×10^{-5}) compared to the full sample (*p*-threshold = 0.01). The size of GxE across sexes at *p*-threshold = 1×10^{-5} were significantly different (GxE*sex *p* = 0.017), but not at the best p-threshold in the full cohort (*p*-threshold = 0.01, GxE*sex *p* = 0.32; **Figure 4.2a**). In women, GxE effect with DSLE predicted depression score (*p*-threshold = 1×10^{-5}; β = 0.039, s.e. = 0.019, R^2 = 0.15%, *p* = 0.038; **Figure 4.2b** and **Appendix C**: Supplementary Figure 1a), as did the GxE effect with ISLE (*p*-threshold = 1×10^{-5}; β = 0.040, s.e. = 0.019, R^2 = 0.16%, *p* = 0.033; **Figure 4.2c** and **Appendix C**: Supplementary Figure 1b). No significant interaction was detected in men (best-fit *p*-threshold = 0.1) with either TSLE (β = 0.039, s.e. = 0.022, R^2 = 0.15%, *p* = 0.072; **Figure 4.2a**), DSLE (β = 0.024, s.e. = 0.022, R^2 = 0.06%, *p* = 0.28; **Figure 4.2b**) or ISLE (β = 0.043, s.e. = 0.022, R^2 = 0.18%, *p* = 0.055; **Figure 4.2c**).

To examine these results further and investigate the *diathesis* effect at different levels of stress, nominally significant GxE were plotted between PRS and categories of SLE (i.e, “none”, “low” and “high” SLE reported; **Figure 4.3**). Examining the interaction found at the full cohort (PRS at PGC-MDD GWAS *p*-threshold = 0.01), we detected a significant direct *diathesis*

effect on the risk of depressive symptoms in those participants reporting SLE, with a higher risk when greater numbers of SLE were reported ("low" number of SLE reported: PRS' β = 0.043, s.e. = 0.021, *p* = 0.039; "high" number of SLE reported: PRS' β = 0.142, s.e. = 0.039, *p* = 2.86 x 10^{-4}; see **Table 4.1** and **Figure 4.3a**). Whereas, in participants who reported no SLE over the preceding 6 months, the risk of depressive symptoms was the same regardless of their *diathesis* risk ("none" SLE reported: PRS' β = 0.021, s.e. = 0.022, *p* = 0.339). Stratifying these results by sex, we found the same pattern as in the full cohort in women ("none": *p* = 0.687; "low": *p* = 0.023; "high": *p* = 2 x 10^{-3}), but not in men ("none": *p* = 0.307; "low": *p* = 0.728; "high": *p* = 0.053; see **Table 4.1** and **Figure 4.3a**). However, the lack of significant *diathesis* effect in men may be due to their lower sample size and its corresponding reduced power.

Examining the interaction with PRS (at PGC-MDD GWAS *p*-threshold = 1 x 10^{-5}) with which a significant interaction was detected in women, we only detected a significant *diathesis* effect on depression score when stratifying by sex in those participants who did not reported SLE over the last 6 months (see **Table 4.1**). The *diathesis* effect was positive in men (PRS' β = 0.082, s.e. = 0.034, *p* = 0.016, R^2 = 0.7%; **Figure 4.3b**), consistent with the contribution of risk alleles. Conversely, the *diathesis* effect was negative in women (PRS' β = -0.061, s.e. = 0.029, *p* = 0.037, R^2 = 0.4%; **Figure 4.3b**), suggesting a protective effect of increasing PRS in those women reporting no SLE, and consistent with the contribution of alleles to individual sensitivity to both positive and negative environmental effects (i.e. "plasticity alleles" rather than "risk alleles")[229,230]. This PRS accounted for the effect of just 34 SNPs, and the size of its GxE across sexes were significantly different (GxE*sex *p* = 0.017; **Figure 4.2a**), supporting possible differences in the underlying stress-response mechanisms between women and men.

Table 4.1 *Diathesis* effect on depression score under SLE categories

Reported values at *p*-thresholds where nominally significant GxE effects were detected.

PRS at *p* value threshold = 0.01									
Sample	***FULL COHORT**			**WOMEN**			**MEN**		
SLE category	**none**	**low**	**high**	**none**	**low**	**high**	**none**	**low**	**high**
N	1833	2311	**775**	1041	1459	490	792	852	285
Effect	0.021	0.043	**0.142**	0.0118	0.0617	0.1538	0.0346	0.0113	0.1227
s.e.	0.022	0.021	**0.039**	0.029	0.027	0.049	0.034	0.032	0.063
t	0.957	2.07	**3.644**	0.403	2.274	3.112	1.021	0.348	1.947
***p* value**	0.339	0.039	**2.86×10^{-4}**	0.687	0.023	0.002	0.307	0.728	0.053
CI (95%)	-0.022, 0.065	0.002, 0.084	**0.065, 0.218**	-0.046, 0.069	0.008, 0.115	0.057, 0.251	-0.032, 0.101	-0.052, 0.075	-0.001,0.247

PRS at *p* value threshold = 1×10^{-5}									
Sample	**FULL COHORT**			***WOMEN**			**MEN**		
SLE category	**none**	**low**	**high**	**none**	**low**	**high**	**none**	**low**	**high**
N	1833	2311	775	1041	1459	490	792	852	285
Effect	-0.0022	0.0032	0.0705	-0.061	0.014	0.078	0.082	-0.0176	0.0548
s.e.	0.022	0.021	0.04	0.029	0.027	0.049	0.034	0.033	0.07
t	-0.098	0.153	1.76	-2.086	0.541	1.609	2.416	-0.537	0.778
***p* value**	0.922	0.878	0.079	0.037	0.589	0.108	0.016	0.592	0.437
CI (95%)	-0.046, 0.041	-0.037, 0.044	-0.008, 0.149	-0.119, -0.004	-0.038, 0.066	-0.017, 0.174	0.015, 0.149	-0.082,0.047	-0.084, 0.193

*Sample where nominally significant GxE was detected. SLE categories (number of SLE reported): 0 SLE = "none", 1 or 2 SLE = "low", and 3 or more SLE = "high". In red, nominally significant effects. In **bold red**, robustly significant effect after conservative Bonferroni correction ($p < 6.94 \times 10^{-4}$).

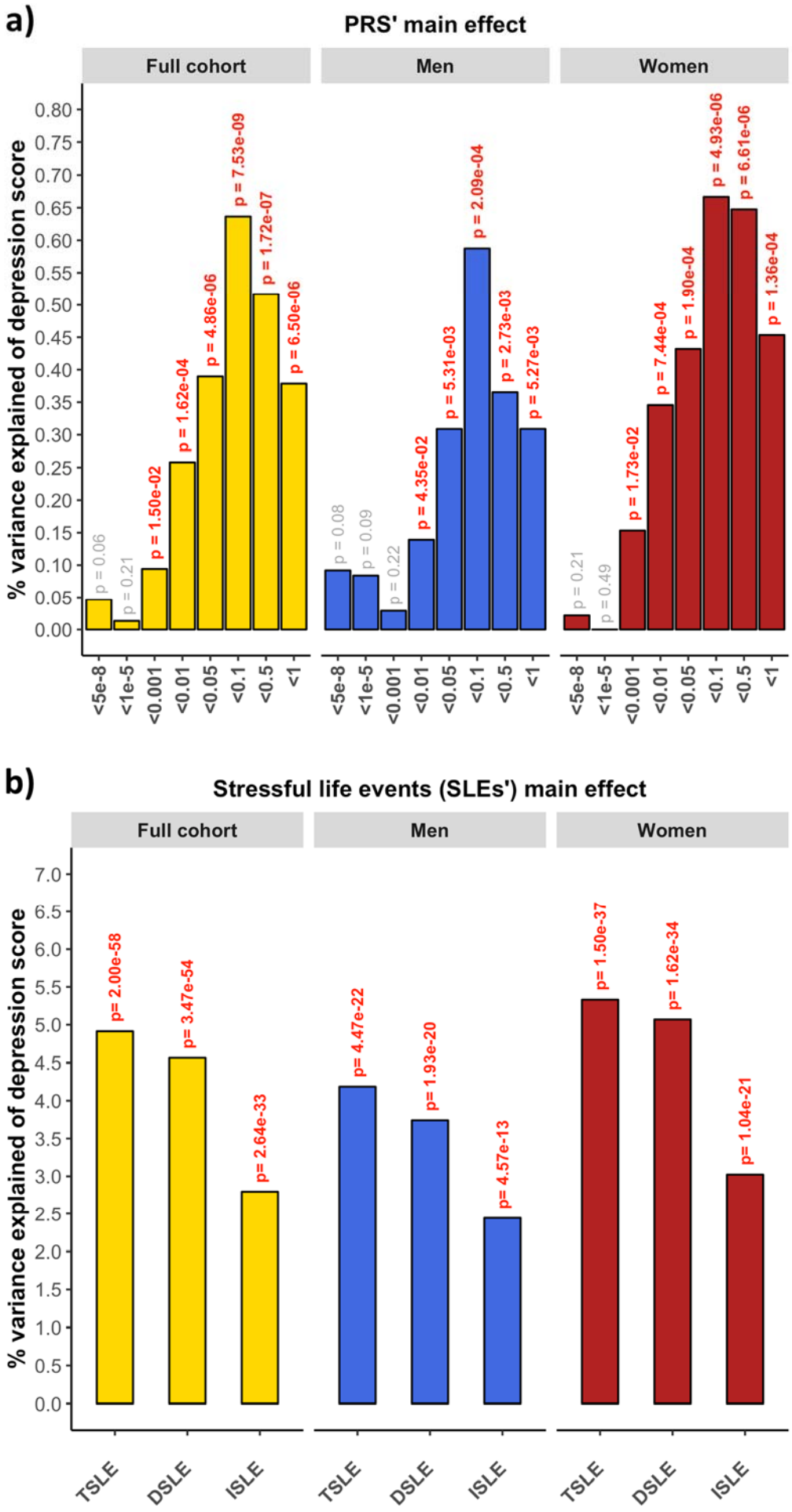

a)
PRS' main effect
Full cohort
Men
Women
% variance explained of depression score
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
p = 0.06
p = 0.21
p = 1.50e-02
p = 1.62e-04
p = 4.86e-06
p = 7.53e-09
p = 1.72e-07
p = 6.50e-06
p = 0.08
p = 0.09
p = 0.22
p = 4.35e-02
p = 5.31e-03
p = 2.09e-04
p = 2.73e-03
p = 5.27e-03
p = 0.21
p = 0.49
p = 1.73e-02
p = 7.44e-04
p = 1.90e-04
p = 4.93e-06
p = 6.61e-06
p = 1.36e-04
<5e-8
<1e-5
<0.001
<0.01
<0.05
<0.1
<0.5
<1
b)
Stressful life events (SLEs') main effect
Full cohort
Men
Women
% variance explained of depression score
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
p= 2.00e-58
p= 3.47e-54
p= 2.64e-33
p= 4.47e-22
p= 1.93e-20
p= 4.57e-13
p= 1.50e-37
p= 1.62e-34
p= 1.04e-21
TSLE
DSLE
ISLE

Figure 4.1 a) Association between polygenic risk scores (PRS) and depression score *(main effects, one-sided tests). PRS were generated at 8 p-threshold levels using summary statistics from the Psychiatric Genetic Consortium MDD GWAS (released July 2016) with the exclusion of Generation Scotland participants. The depression score was derived from The General Health Questionnaire. The Y-axis represents the % of variance of depression score explained by PRS main effects. The full cohort (yellow) was split into men (blue) and women (red). In Colodro-Conde et al., PRS for MDD significantly explained up to 0.46% of depression score in their sample (~0.39% in women and ~0.70% in men).* **b) Association between reported number of SLE and depression score** *(main effect, one-sided tests, results expressed in % of depression score explained). SLE were self-reported through a brief life-events questionnaire based on the List of Threatening Experiences and categorized into: total number of SLE reported (TSLE), dependent SLE (DSLE) or independent SLE (ISLE). The full cohort (yellow) was split into men (blue) and women (red). In Colodro-Conde et al., "personal" SLE significantly explained up to 12.9% of depression score variance in their sample (~11.5% in women and ~16% in men).*

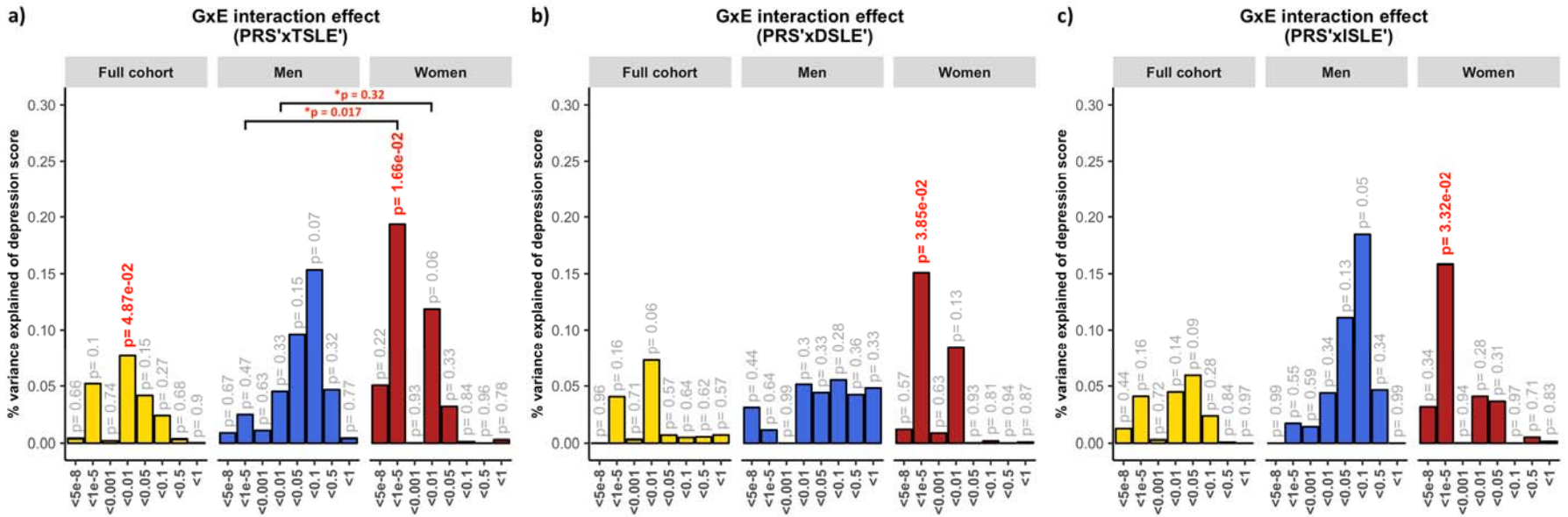


Figure 4.2 Association between GxE effect and depression score. *The results represent percentage of depression score explained by the interaction term (two-sided tests) fitted in linear mixed models to empirically test the diathesis-stress model. Red numbers show significant interactions p-values. *Shows significance of difference between sexes when comparing the size of the estimated GxE effects. The full cohort (yellow) was split into men (blue) and women (red). PRS were generated at 8 p-threshold levels using summary statistics from the Psychiatric Genetic Consortium MDD GWAS (released July 2016) with the exclusion of Generation Scotland participants. The interaction effect was tested with **a)** the number of SLE reported (TSLE), **b)** dependent SLE (DSLE) and **c)** independent SLE (ISLE). In Colodro-Conde et al., the variance of depression score explained in their sample by GxE was 0.12% ($p = 7 \times 10^{-3}$). GxE were also significant in women ($p = 2 \times 10^{-3}$) explaining up to 0.25% of depression score variation, but not in men ($p = 0.059$; $R^2 = 0.17\%$; negative/protective effect on depression score).*

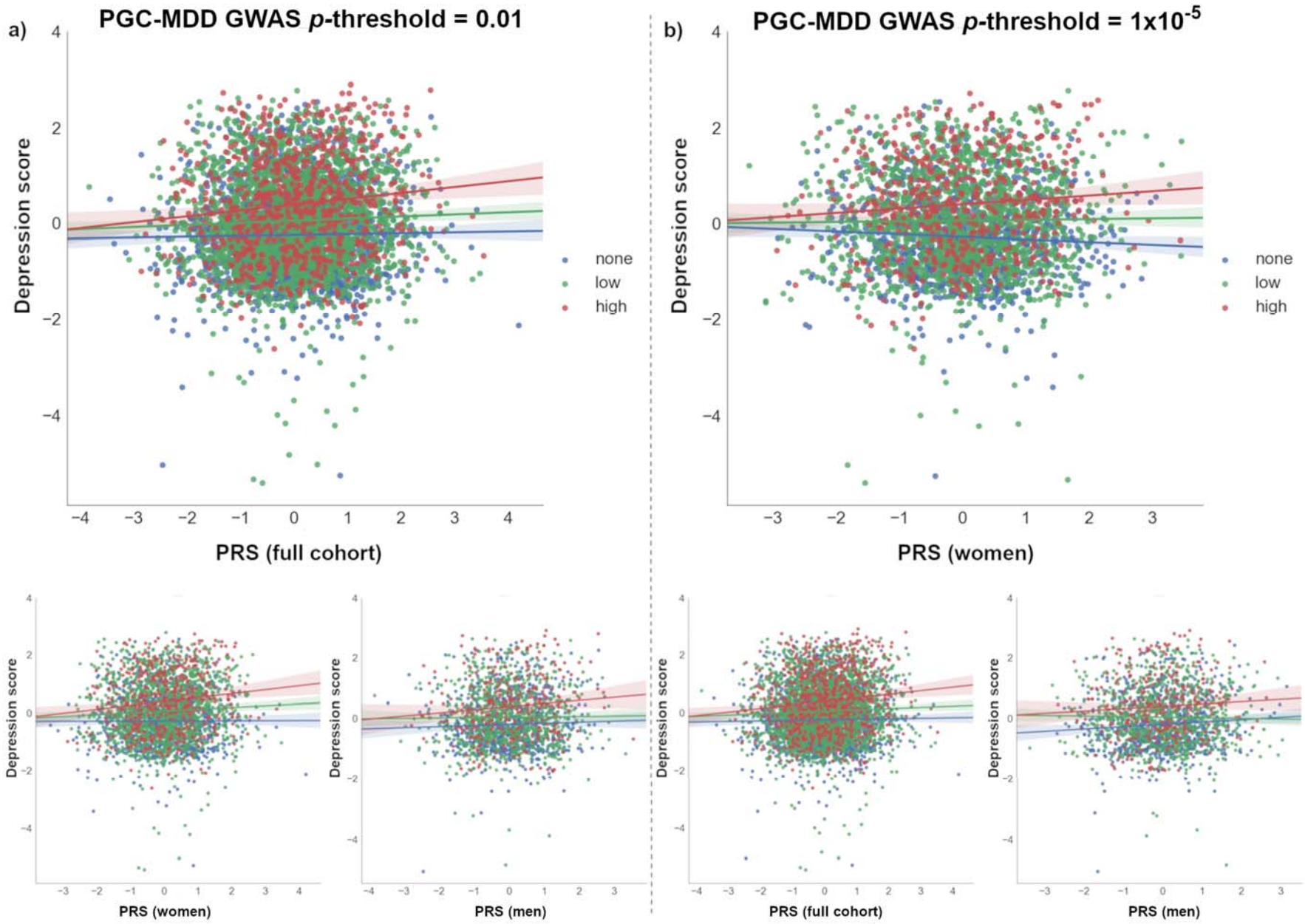

a)
PGC-MDD GWAS p-threshold = 0.01
Depression score
PRS (full cohort)
none
low
high
b)
PGC-MDD GWAS p-threshold = 1x10-5
Depression score
PRS (women)
none
low
high
Depression score
PRS (women)
Depression score
PRS (men)
Depression score
PRS (full cohort)
Depression score
PRS (men)

Figure 4.3 Scatterplot of *diathesis-stress* interactions on depression score. *Interactions with PRS at which nominally significant GxE effects were detected in **a)** full cohort (p-threshold = 0.01) and **b)** in women (p-threshold = 1 x 10^{-5}) are shown. At bottom, the remaining samples (i.e., full cohort, women or men) at same p-threshold are shown for comparison. The X-axis represents the direct effect of PRS (standard deviation from the mean) based on **a)** p-threshold = 0.01 and **b)** p-threshold = 1 x 10^{-5}, using the total number of SLE reported by each participant (dot) as environmental exposures at three SLE levels represented by colours. Blue: 0 SLE, "no stress", n = 1,833/1,041/792; green: 1 or 2 SLE, "low stress", n = 2,311/1,459/852; red: 3 or more SLE, "high stress", n = 775/490/285; in the full cohort, women and men, respectively. Y-axis reflects the depression score standardized to mean of 0 and standard deviation of 1. Lines represent the increment in risk of depression under a certain degree of "stress" dependent on a genetic predisposition (= diathesis).*

4.5 Discussion

The findings reported in this study support those from Colodro-Conde *et al*.[169], in an independent sample of similar sample size and study design, and also support possible sex-specific differences in the effect of genetic risk of MDD in response to SLE.

Both Colodro-Conde *et al.* and our study suggest that individuals with an inherent genetic predisposition to MDD, reporting high number of recent SLE, are at additional risk of depressive symptoms due to GxE effects, thus validating the *diathesis-stress* theory. We identified nominally significant GxE effects in liability to depression at the population level (p = 0.049) and in women (p = 0.017), but not in men (p = 0.072). However, these interactions did not survive multiple testing correction ($p > 2.08 \times 10^{-3}$) and the power of both studies to draw robust conclusions remains limited[170]. With increased power these studies could determine more accurately both the presence and magnitude of a GxE effect in depression. To better understand the effect of PRS at different levels of exposure to stress, we examined the nominally significant interactions detected in the full sample by categorizing participants on three groups based on the number of SLE reported (i.e. “none”, “low” or “high”). We detected a significant *diathesis* effect on risk of depression only in those participants reporting SLE, but not in those participants that reported no SLE over the last 6 month. Furthermore, the *diathesis* effect was stronger on those participants reporting a “high” number of SLE (β = 0.142, $p = 2.86 \times 10^{-4}$) compared to those participants reporting a “low” number of SLE (β = 0.043, p = 0.039). The former effect was robustly significant and survived a conservative Bonferroni correction to adjust for multiple testing ($p < 6.94 \times 10^{-4}$). This finding corroborates the *diathesis-stress* model for depression and supports Colodro-Conde *et al*. results using an independent sample.

To investigate the relative contribution of the GxE to the variance of depression, we examined in the full cohort the total variance of depression

score explained by the PRS main effect and the significant GxE effect jointly. Together, they explained 0.34% of the variance, of which 0.07% of the variance of the depression score was attributed to the GxE effect (*p*-threshold = 0.01; PRS $p = 1.19 \times 10^{-4}$, GxE $p = 0.049$; both derived from the full diathesis-model with TSLE). This is lower than the proportion of variance attributed to common SNPs (8.9%) in the full PGC-MDD analysis[150]. As Colodro-Conde *et al.* noted, this result aligns with estimates from experimental organisms suggesting that around 20% of the heritability may be typically attributed to the effects of GxE[425], although it is inconsistent with the majority of human traits with the potential exception of depression[82].

Consistent with PRS predicting "personal" SLE in Colodro-Conde *et al.,* PRS for MDD predicted SLE in our study (see **Appendix C**: Supplementary Figure 2), although not at the *p*-threshold at which significant GxE effects were detected. Genetic factors predisposing to MDD may contribute to individuals exposing themselves to, or showing an increased reporting of, SLE via behavioural or personality traits[79,98]. Such genetic mediation of the association between depression and SLE would disclose a gene-environment correlation (i.e. genetic effects on the probability of undergoing a SLE) that hinders to interpret our findings as pure GxE effects[74,426]. To address this limitation and assess this aspect, following Colodro-Conde *et al*., we split the 12-items TSLE measure into SLE that are either potentially "dependent" on a participant's own behaviour (DSLE; therefore, potentially driven by genetic factors) or not ("independent" SLE; ISLE)[99,100]. DSLE are reported to be more heritable and have stronger associations with MDD than ISLE[98,100,109]. In our sample, reporting DSLE is significantly heritable ($h^2_{SNP} = 0.131$, s.e. = 0.071, $p = 0.029$), supporting a genetic mediation of the association, whereas reporting ISLE is not significantly heritable ($h^2_{SNP} = 0.000$, s.e. = 0.072, $p = 0.5$)[427]. Nominally significant GxE effects were seen in women for both DSLE and ISLE, suggesting that both GxE and gene-environment correlation co-occur. Colodro-Conde *et al*. did not identify significant GxE using independent SLE as the exposure.

Between-sex differences on stress response could help to explain previous differences seen between sexes in depression such as those in associated risk (i.e. approximately 1.5 - 2-fold higher in women), symptoms reported and/or coping strategies (e.g., whereas women tend to cope through verbal and emotional strategies, men tend to cope by doing sport and consuming alcohol)[49-53]. This also aligns with an increased risk associated with a lack of social support seen in women compared to men[169]. Furthermore, although we do not know whether participants experienced recent events with positive effects, we saw a protective effect in those women who did not experienced recent SLE (p = 0.037), suggesting that some genetic variants associated with MDD may operate as "plasticity alleles" and not just as "risk alleles"[229,230]. This effect was neutralized in the full cohort due to an opposite effect in men (p = 0.016), but it is supported by previous protective effects reported when using a serotoninergic multilocus profile score and absence of SLE in young women[428]. These findings would be consistent with a differential-susceptibility model[226,227] of depression, also suggested by the interaction effects seen between the serotonin transporter linked promoter region gene (5-HTTLPR) locus and family support and liability to adolescent depression in boys[429]. However, our results and the examples given are only nominally significant and will require replication in larger samples. Robustly identified sex-specific differences in genetic stress-response could improve personalized treatments and therapies such as better coping strategies.

There are notable differences between our study and Colodro-Conde *et al.* to consider before accepting our findings as a replication of Colodro-Conde *et al.* results. First, differences in PRS profiling may have affected replication power. We used the same equivalent PGC-MDD2 GWAS as discovery sample. However, whereas Colodro-Conde *et al.* generated PRS in their target sample containing over 9.5M imputed SNP, in this study we generated PRS in a target sample of over 560K genotyped SNPs (see **Appendix C**: Supplementary Table 1 for comparison). This potentially results in a less informative PRS in our study, with less predictive power, although the variance explained by our PRS was slightly larger (0.64% vs. 0.46%). The

size of the discovery sample is key to constructing an accurate predictive PRS, but to exploit the most of the variants available may be an asset[170]. Secondly, different screening tools were used to measure both current depression and recent environmental stressors across the two studies. Both studies transformed their data, using item response theory or by log-transformation, to improve the data distribution. However, neither study used depression scores that were normally distributed. The scale of the instruments used and their corresponding parameterization to test an interaction could have a direct effect on the size and significance of their interaction[425,430]; so findings from GxE must be taken with caution. Furthermore, although both screening methods have been validated and applied to detect depressive symptoms, different questions may cover and emphasise different features of the illness, which may result in different outputs. The same applies to the measurement of environmental stressors in the two studies. Both covering of a longer time-period and upweighting by *dependent* SLE items may explain the increased explanatory power of "personal" SLE (12.9%) in Colodro-Conde *et al*. to predict depression score compared to our "total" SLE measure (4.91%). Finally, the unmeasured aspects of the exposure to SLE or its impact may also contribute to lack of stronger replication and positive findings.

In conclusion, despite differences in the measures used across studies, we saw concordance and similar patterns between our results and those of Colodro-Conde *et al.*[169]. Our findings are consistent with Colodro-Conde *et al.* and, therefore, add validity to the *diathesis-stress* theory for depression. Empirically demonstrating the *diathesis-stress* theory for depression would validate recent[113,267-269] and future studies using a genome-wide approach to identify genetic mechanisms and interactive pathways involved in GxE underpinning the causative effect of "stress" in the development of depressive symptoms and mental illness in general. This study adds to our understanding of gene-by-environment interactions, although larger samples will be required to confirm differences in *diathesis-stress* effects between women and men.

Chapter 5 Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland

As seen in **chapter 4**, I detected a significant PRSxE effect on depression score supporting, under a *diathesis-stress* model, the presence of GxE effects underlying MDD. My findings are in accordance with Colodro-Conde *et al.* results and, together, validate the implementation of hypothesis-free GxE studies at the SNP level through genome-wide by environment interaction studies (GWEIS). Therefore, whereas in **chapter 4** I sought interactions at the individual level using PRS as an individual's vulnerability score, in this chapter I seek interactions at the SNP level by performing GWEIS in the subsample from Generation Scotland, and a second sample from UK Biobank, using measures for the construction of SLE reported and depression score.

This chapter is published in *Translational Psychiatry* and is shown as it has been accepted, which explains the use of "we" within the chapter. I confirm that the work of this chapter is my own work under guidance from my supervisor Dr. Pippa Thomson. I carried out all the analyses myself. **Appendix D** includes the published article and all **Supplementary Material**.

Publication:

Arnau-Soler, A. *et al.* Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational Psychiatry* 9, 14 (2019).

5.1 Abstract

Stress is associated with poorer physical and mental health. To improve our understanding of this link, we performed genome-wide association studies (GWAS) of depressive symptoms and genome-wide by environment interaction studies (GWEIS) of depressive symptoms and stressful life events (SLE) in two UK population cohorts (Generation Scotland and UK Biobank). No SNP was individually significant in either GWAS, but gene-based tests identified six genes associated with depressive symptoms in UK Biobank (*DCC, ACSS3, DRD2, STAG1, FOXP2 and KYNU; p* < 2.77 x 10^{-6}). Two SNPs with genome-wide significant GxE effects were identified by GWEIS in Generation Scotland: rs12789145 (53kb downstream *PIWIL4*; *p* = 4.95 x 10^{-9}; total SLE) and rs17070072 (intronic to *ZCCHC2*; *p* = 1.46 x 10^{-8}; dependent SLE). A third locus upstream *CYLC2* (rs12000047 and rs12005200, *p* < 2.00 x 10^{-8}; dependent SLE) when the joint effect of the SNP main and GxE effects was considered. GWEIS gene-based tests identified: *MTNR1B* with GxE effect with dependent SLE in Generation Scotland; and *PHF2* with the joint effect in UK Biobank (*p* < 2.77 x 10^{-6}). Polygenic risk scores (PRS) analyses incorporating GxE effects improved the prediction of depressive symptom scores, when using weights derived from either the UK Biobank GWAS of depressive symptoms (*p* = 0.01) or the PGC GWAS of major depressive disorder (*p* = 5.91 x 10^{-3}). Using an independent sample, PRS derived using GWEIS GxE effects provided evidence of shared aetiologies between depressive symptoms and schizotypal personality, heart disease and COPD. Further such studies are required and may result in improved treatments for depression and other stress-related conditions.

5.2 Introduction

Mental illness results from the interplay between genetic susceptibility and environmental risk factors[12,74]. Previous studies have shown that the effects of environmental factors on traits may be partially heritable[97] and moderated by genetics[114,169]. Major depressive disorder (MDD) is the most common psychiatric disorder with a lifetime prevalence of approximately 14% globally[9] and with a heritability of approximately 37%[119]. There is strong evidence for the role of stressful life events (SLE) as risk factor and trigger for depression[84,90,93,95,401]. Genetic control of sensitivity to stress may vary between individuals, resulting in individual differences in the depressogenic effects of SLE, i.e., genotype-by-environment interaction (GxE)[103,115,169,378,431]. Significant evidence of GxE has been reported for common respiratory diseases and some forms of cancer[432-437], and GxE studies have identified genetic risk variants not found by genome-wide association studies (GWAS)[438-442].

Interaction between polygenic risk of MDD and recent SLE are reported to increase liability to depressive symptoms[169,431]; validating the implementation of genome-wide approaches to study GxE in depression. Most GxE studies for MDD have been conducted on candidate genes, or using polygenic approaches to a wide range of environmental risk factors, with some contradictory findings[252,253,275,290,443]. Incorporating knowledge about recent SLE into GWAS may improve our ability to detect risk variants in depression otherwise missed in GWAS[233]. To date, four studies have performed genome-wide by environment interaction studies (GWEIS) of MDD and SLE[113,267-269], but this is the first study to perform GWEIS of depressive symptoms using adult SLE in cohorts of relatively homogeneous European ancestry.

Interpretation of GxE effects may be hindered by gene-environment correlation. Gene-environment correlation denotes a genetic mediation of

associations through genetic influences on exposure to, or reporting of, environments[74,426]. Genetic factors predisposing to MDD may contribute to exposure and/or reporting of SLE[98]. To tackle this limitation, measures of SLE can be broken down into SLE likely to be independent of a respondent's own behaviour and symptoms, or into dependent SLE, in which participants may played an active role exposure to SLE[99,100]. Different genetic influences with a higher heritability for reported dependent SLE than independent SLE[98,109-112] suggest that whereas GxE driven by independent SLE is likely to reflect a genetic moderation of associations between SLE and depression, GxE driven by dependent SLE may result from a genetic mediation of the association through genetically driven personality or behavioural traits. To test this we analysed dependent and independent SLE scores separately in Generation Scotland.

Stress contributes to many human conditions, with evidence of genetic vulnerability to the effect of SLE[444]. Therefore, genetic stress-response factors in MDD may also underlie the aetiology of other stress-linked disorders, with which MDD is often co-morbid[3,4] (e.g. cardiovascular diseases[66], diabetes,[67] chronic pain[68] and inflamation[69]). We tested the hypothesis that pleiotropy and shared aetiology between mental and physical health conditions may be due in part to genetic variants underlying SLE effects in depression.

In this study we conduct GWEIS of depressive symptoms incorporating data on SLE in two independent UK-based cohorts. We aimed to: i) identify loci associated with depressive symptoms through genetic response to SLE; ii) study dependent and independent SLE to support a contribution of genetically mediated exposure to stress; iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone; and iv) test for a significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes.

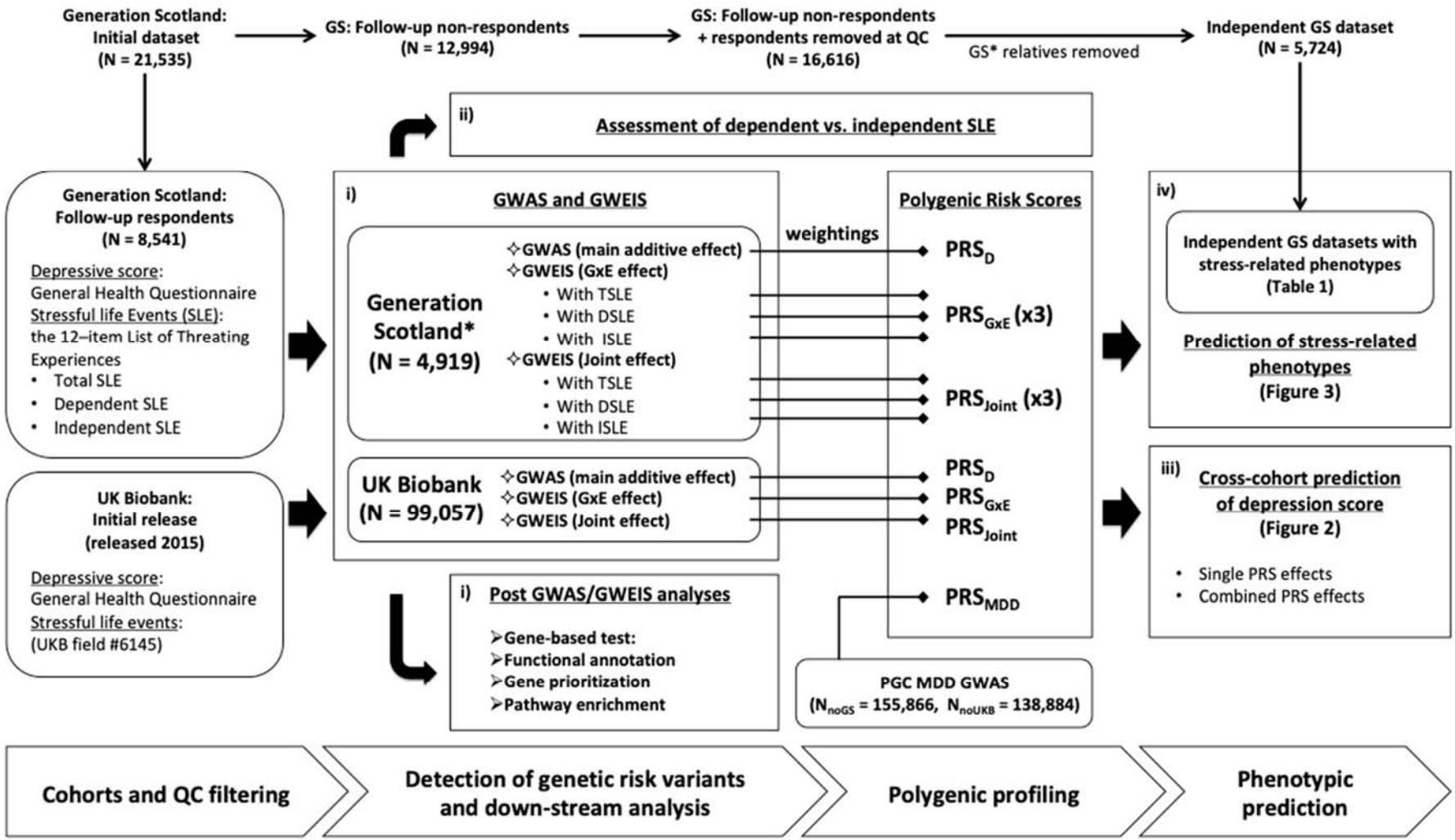

Generation Scotland: Initial dataset (N = 21,535)
GS: Follow-up non-respondents (N = 12,994)
GS: Follow-up non-respondents + respondents removed at QC (N = 16,616)
GS* relatives removed
Independent GS dataset (N = 5,724)
ii) Assessment of dependent vs. independent SLE
Generation Scotland: Follow-up respondents (N = 8,541)
Depressive score: General Health Questionnaire
Stressful life Events (SLE): the 12–item List of Threating Experiences
• Total SLE
• Dependent SLE
• Independent SLE
UK Biobank: Initial release (released 2015)
Depressive score: General Health Questionnaire
Stressful life events: (UKB field #6145)
i) GWAS and GWEIS
Generation Scotland* (N = 4,919)
✧GWAS (main additive effect)
✧GWEIS (GxE effect)
• With TSLE
• With DSLE
• With ISLE
✧GWEIS (Joint effect)
• With TSLE
• With DSLE
• With ISLE
UK Biobank (N = 99,057)
✧GWAS (main additive effect)
✧GWEIS (GxE effect)
✧GWEIS (Joint effect)
weightings
Polygenic Risk Scores
PRS$_D$
PRS$_{GxE}$ (x3)
PRS$_{Joint}$ (x3)
PRS$_D$
PRS$_{GxE}$
PRS$_{Joint}$
PRS$_{MDD}$
i) Post GWAS/GWEIS analyses
➢Gene-based test:
➢Functional annotation
➢Gene prioritization
➢Pathway enrichment
PGC MDD GWAS (N_{noGS} = 155,866, N_{noUKB} = 138,884)
iv) Independent GS datasets with stress-related phenotypes (Table 1)
Prediction of stress-related phenotypes (Figure 3)
iii) Cross-cohort prediction of depression score (Figure 2)
• Single PRS effects
• Combined PRS effects
Cohorts and QC filtering
Detection of genetic risk variants and down-stream analysis
Polygenic profiling
Phenotypic prediction

Figure 5.1 Study flowchart. Overview of analysis conducted on this study to achieve our aims: i) identify loci associated with depressive symptoms through genetic response to SLE; ii) test whether results of studying dependent and independent SLE support a contribution of genetically mediated exposure to stress; iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone and iv) test whether there is significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes. Two core cohorts are used, Generation Scotland (GS) and UK Biobank (UKB). Summary statistics from Genome-Wide Association Studies (GWAS) and Genome-Wide by Environment Interaction Studies (GWEIS) are used to generate Polygenic Risk Scores (PRS). Summary statistics from Psychiatric Genetic Consortium (PGC) Major Depressive Disorder (MDD) GWAS are also used to generate PRS (PRS_{MDD}). PRS weighted by: additive effects (PRS_{D} and PRS_{MDD}), GxE effects (PRS_{GxE}) and joint effects (the combined additive and GxE effect; PRS_{Joint}), are used for phenotypic prediction. TSLE stands for Total number of SLE reported. DSLE stands for SLE dependent on an individual's own behaviour. Conversely, ISLE stands for independent SLE. N stands for sample size. N_{noGS} stands for sample size with GS individuals removed. N_{noUKB} stands for sample size with UKB individuals removed.

5.3 Materials and methods

The core workflow of this study is summarized at **Figure 5.1**.

5.3.1 Cohort descriptions

5.3.1.1 Generation Scotland (GS)

Generation Scotland is a family-based population cohort representative of the Scottish population[305]. At baseline, blood and salivary DNA samples were collected, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. Genome-wide genotype data was generated using the Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry[413]. The procedures and details for DNA extraction and genotyping have been extensively described elsewhere[307,414]. 21,525 participants were re-contacted to participate in a follow-up mental health study (Stratifying Resilience and Depression Longitudinally, STRADL), of which 8,541 participants responded providing updated measures in psychiatric symptoms and SLE through self-reported mental health questionnaires[415]. Samples were excluded if: they were duplicate samples, had diagnoses of bipolar disorder, no SLE data (non-respondents), were population outliers (mainly non-Caucasians and Italian ancestry subgroup), had sex mismatches, or were missing more than 2% of genotypes. SNPs were excluded if: missing more than 2% of genotypes, Hardy-Weinberg Equilibrium test $p < 1 \times 10^{-6}$, or minor allele frequency less than 1%. Further details of the GS and STRADL cohort are available elsewhere[124,305,306,415]. All components of GS and STRADL obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the NHS (reference 05/s1401/89). After quality control, individuals were filtered by degree of relatedness (pi-hat < 0.05), maximizing retention of those individuals reporting a higher number of SLE. The final dataset comprised data on 4,919 unrelated individuals (1,929 men; 2,990 women) and 560,351 SNPs.

5.3.1.2 Independent GS datasets

Additional datasets for a range of stress-linked medical conditions and personality traits were created from GS (N = 21,525) excluding respondents and their relatives (N = 5,724). Following the same quality control criteria detailed above, we maximized unrelated non-respondents for retention of cases, or proxy cases (see below), to maximize the information available for each phenotype. This resulted in independent datasets with unrelated individuals for each trait. Differences between respondents and non-respondents are noted in the legend of **Table 5.1**.

5.3.1.3 UK Biobank (UKB)

This study used data from 99,057 unrelated individuals (47,558 men; 51,499 women) from the initial release of UKB genotyped data (released 2015; under UK Biobank project 4844). Briefly, participants were removed based on UKB genomic analysis exclusion, non-white British ancestry, high missingness, genetic relatedness (kinship coefficient > 0.0442), QC failure in UK BiLEVE study, and gender mismatch. GS participants and their relatives were excluded and GS SNPs imputed to a reference set combining the UK10K haplotype and 1000 Genomes Phase 3 reference panels[445]. After quality control, 1,009,208 SNPs remained. UK Biobank received ethical approval from the NHS National Research Ethics Service North West (reference: 11/NW/0382). Further details on UKB cohort description, genotyping, imputation and quality control are available elsewhere[301,302,446].

All participants provided informed consent.

5.3.2 Phenotype assessment

5.3.2.1 Stressful life events (SLE)

GS participants reported SLE experienced over the preceding 6 months through a self-reported brief life events questionnaire based on the 12–item List of Threatening Experiences[99,422,423] (**Appendix D**: Supplementary Table 1a). The total number of SLE reported (TSLE) consisted of the number of 'yes' responses. TSLE were subdivided into SLE potentially dependent or secondary to an individual's own behaviour (DSLE, questions 6-11 in

Appendix D: Supplementary Table 1a), and independent SLE (ISLE, questions 1-5 in **Appendix D**: Supplementary Table 1a; pregnancy item removed) following Brugha *et al.*[99,100]. Thus, 3 SLE measures (TSLE, DSLE and ISLE) were constructed for GS. UKB participants were screened for "*illness, injury, bereavement and stress*" (**Appendix D**: Supplementary Table 1b) over the previous 2 years using 6 items included in the UKB Touchscreen questionnaire. A score reflecting SLE reported in UKB ($TSLE_{UKB}$) was constructed by summing the number of 'yes' responses.

5.3.2.2 Psychological assessment

GS participants reported whether their current mental state over the preceding 2 weeks differed from their typical state using a self-administered 28-item scaled version of The General Health Questionnaire (GHQ)[197,416,417]. Participants rated the degree and severity of their current symptoms with a four-point Likert scale (following Goldberg et al., 1997[197]). A final log-transformed GHQ was used to detect altered psychopathology and thus, assess depressive symptoms as results of SLE. In UKB participants, current depressive symptoms over the preceding 2 weeks were evaluated using 4 psychometric screening items (**Appendix D**: Supplementary Table 2), including two validated and reliable questions for screening depression[447], from the Patient Health Questionnaire (PHQ) validated to screen mental illness[310,311]. Each question was rated in a four-point Likert scale to assess impairment/severity of symptoms. Due to its skewed distribution, a four-point PHQ score was formed from PHQ (0 = 0; 1 = 1-2; 2 = 3-5; 3 = 6 or more) to create a more normal distribution.

5.3.2.3 Stress-related traits

Targeted GS stress-related phenotypes and sample sizes are shown in **Table 5.1** and detailed elsewhere[305]. These conditions were selected from literature review based on previous evidence of a link with stress[444] (see also **Appendix D.3**). Furthermore, we created additional independent samples using mapping by proxy, where individuals with a self-reported first-degree relative with a selected phenotype were included as proxy cases. This

approach provides greater power to detect susceptibility variants in traits with low prevalence[448].

5.3.3 Statistical analyses

5.3.3.1 SNP-heritability and genetic correlation

Restricted maximum likelihood approach was applied to estimate SNP-heritability (h^2_{SNP}) of depressive symptoms and self-reported SLE measures, and within samples bivariate genetic correlation between depressive symptoms and SLE measures using GCTA[176].

5.3.3.2 GWAS analyses

GWAS were conducted in PLINK[273]. In GS, age, sex and 20 principal components (PCs) were fitted as covariates. In UKB, age, sex, and 15 PCs recommended by UKB were fitted as covariates. The genome-wide significance threshold was $p = 5 \times 10^{-8}$.

5.3.3.3 GWEIS analyses

GWEIS were conducted on GHQ (the dependent variable) for TSLE, DSLE and ISLE in GS and on PHQ for $TSLE_{UKB}$ in UKB fitting the same covariates detailed above to reduce error variance. GWEIS were conducted using an R plugin for PLINK[273] developed by Almli *et al.*[272] (https://epstein-software.github.io/robust-joint-interaction). This method implements a robust test, that jointly considers SNP and SNP-environment interaction effects from a full model ($Y \sim \beta_0 + \beta SNP + \beta SLE + \beta SNPxSLE + \beta Covariates$) against a null model where both the SNP and SNPxSLE effects equal 0, to assess the joint effect (the combined additive main and GxE genetic effect at a SNP) using a nonlinear statistical approach that applies Huber-White estimates of variance to correct possible inflation due to heteroscedasticity (unequal variances across exposure levels). This robust test should reduce confounding due to differences in variance induced by covariate interaction effects[424] if present. Additional code was added (courtesy of Prof. Michael Epstein[272]; **Appendix D.1**) to generate beta-coefficients and the *p*-value of the GxE term alone. In UKB, correcting for 1,009,208 SNPs and 1 exposure we established a Bonferroni-adjusted threshold for significance at $p = 2.47 \times$

10^{-8} for both joint and GxE effects. In GS, correcting for 560,351 SNPs and 3 measures of SLE we established a genome-wide significance threshold of $p = 2.97 \times 10^{-8}$.

5.3.3.4 Post-GWAS/GWEIS analyses

GWAS and GWEIS summary statistics were analysed using FUMA[316] including: gene-based tests, functional annotation, gene prioritization and pathway enrichment (**Appendix D.2**).

5.3.3.5 Polygenic profiling & prediction

Polygenic risk scores (PRS) weighting by GxE effects (PRS_{GxE}) were generated using PRSice-2[313] (**Appendix D.2**) in GS using GxE effects from UKB-GWEIS. In UKB, PRS_{GxE} were constructed using GxE effects from all three GS-GWEIS (TSLE, DSLE and ISLE as exposures) independently. PRS were also weighted in both samples using either UKB-GWAS or GS-GWAS statistics (PRS_D), and summary statistics from Psychiatric Genetic Consortium (PGC) MDD-GWAS (released 2016; PRS_{MDD}) that excluded GS and UKB individuals when required (N_{noGS} = 155,866; N_{noUKB} = 138,884). Furthermore, we calculated PRS weighted by the joint effects (the combined additive main and GxE genetic effects; PRS_{Joint}) from either the UKB-GWEIS or GS-GWEIS. PRS predictions of depressive symptoms were permuted 10,000 times. Multiple regression models fitting PRS_{GxE} and PRS_{MDD}, and both PRS_{GxE} and PRS_D were tested. All models were adjusted by same covariates used in GWAS/GWEIS. Null models were estimated from the direct effects of covariates alone. The predictive improvement of combining PRS_{GxE} and PRS_{MDD}/PRS_D effects over PRS_{MDD}/PRS_D effect alone was tested for significance using the likelihood-ratio test (LRT).

Prediction of PRS_D, PRS_{GxE} and PRS_{Joint} on stress-linked traits were adjusted by age, sex and 20 PCs; and permuted 10,000 times. Empirical-*p*-values after permutations were further adjusted by false discovery rate (conservative threshold at *Empirical-p* = 6.16×10^{-3}). The predictive improvement of fitting PRS_{GxE} combined with PRS_D and covariates over prediction of a phenotype using the PRS_D effect alone with covariates was assessed using LRT, and

LRT-p-values adjusted by FDR (conservative threshold at *LRT-p* = 8.35 x 10^{-4}).

Table 5.1 GS samples with stress-related phenotypes

Trait	N	Males/Females	N SNPs	N Cases	N Controls
Alzheimer (R)	3377	1475/1903	560622	**655**	**2722**
Asthma	3390	1500/1890	560569	**555**	**2835**
Asthma (R)	3375	1470/1905	560432	**910**	**2465**
Bowel cancer (R)	3386	1495/1891	560630	**672**	**2714**
Breast cancer	3388	1486/1902	560611	**83**	**3305**
Breast cancer (R)	3386	1482/1904	560579	**564**	**2822**
Chronic obstructive pulmonary disease	3387	1496/1891	560591	**73**	**3314**
Chronic obstructive pulmonary disease (R)	3387	1474/1913	560620	**553**	**2834**
Depression	3385	1495/1890	560584	**483**	**2902**
Depression (R)	3382	1506/1876	560514	**731**	**2651**
Diabetes	3388	1497/1891	560469	**185**	**3203**
Diabetes (R)	3389	1481/1908	560584	**1144**	**2245**
Heart disease	3392	1504/1888	560526	**212**	**3180**
Heart disease (R)	3377	1483/1894	560479	**2254**	**1123**
High blood preasure	3402	1501/1901	560508	**729**	**2673**
High blood preasure (R)	3372	1464/1908	560569	**1901**	**1471**
Hip fracture (R)	3388	1489/1899	560572	**421**	**2967**
Lung cancer (R)	3379	1492/1887	560600	**798**	**2581**
Osteoarthritis	3395	1486/1909	560640	**411**	2984
Osteoarthritis (R)	3383	1466/1917	560516	**961**	**2422**
Parkinson (R)	3388	1488/1900	560590	**236**	**3152**
Prostate cancer (R)	3381	1495/1886	560570	**329**	**3052**
Rheumatoid arthritis	3387	1490/1897	560618	**93**	**3294**
Rheumatoid arthritis (R)	3380	1487/1893	560543	**765**	**2615**
Stroke	3387	1492/1895	560613	**81**	**3306**
Stroke (R)	3385	1463/1922	560478	**1506**	**1879**
Neuroticism*	3421	1521/1900	560484	-	-
Extraversion*	3420	1520/1900	560476	-	-
Schizotypal personality*	2386	1065/1321	560369	-	-
Mood disorder*	2307	1040/1267	560318	-	-

Samples were maximized for retention of cases to maximize the information available for each trait. There was no preferential selection of relatives in pairs for quantitative phenotypes, in order to retain the underlying distribution. All individuals involved in the datasets listed above were non-respondents to the GS follow-up study. Compared to individuals included at GS GWEIS (respondents in GS follow-up), non-respondents were significantly: younger, from more socioeconomically deprived areas, generally less healthier and wealthier. Non-respondents were more likely to smoke, and less likely to drink alcohol, although they consumed more units per week, compared with respondents. At GS baseline, non-respondents experienced more psychological distress and reported higher scores in symptoms of GHQ-depression and GHQ-anxiety than respondents[415]. The total target sample size (N), number of males and females in N, number of SNPs (N SNPs) in target sample size N: the number of SNPs used as predictors after clumping step range between 90650 - 91000. The number of cases and controls in the independent target sample is indicated for binary phenotypes only. Samples were mapping by proxy approach was used (i.e. where first-degree relatives of individuals with the disease were considered proxy cases and included into the group of cases) are indicated by (R). *Assessed through self-reported questionnaires.

5.4 Results

5.4.1 Phenotypic and genetic correlations

Depressive symptoms scores and SLE measures were positively correlated in both UKB (r^2 = 0.22, $p < 2.2 \times 10^{-16}$) and GS (TSLE-r^2 = 0.21, $p = 1.69 \times 10^{-52}$; DSLE-r^2 = 0.21, $p = 8.59 \times 10^{-51}$; ISLE-r^2 = 0.17, $p = 2.33 \times 10^{-33}$). Significant bivariate genetic correlation between depression and SLE scores was identified in UKB (rG = 0.72; $p < 1 \times 10^{-5}$, N = 50,000), but not in GS (rG = 1, $p \geq 0.056$, N = 4,919; **Appendix D**: Supplementary Table 3a).

5.4.2 SNP-heritability (h^2_{SNP})

In UKB, a significant h^2_{SNP} of PHQ was identified (h^2_{SNP} = 0.090; $p < 0.001$; N = 99,057). This estimate remained significant after adjusting by $TSLE_{UKB}$ effect (h^2_{SNP} = 0.079; $p < 0.001$), suggesting a genetic contribution unique of depressive symptoms. The h^2_{SNP} of $TSLE_{UKB}$ was also significant (h^2_{SNP} = 0.040, $p < 0.001$; **Appendix D**: Supplementary Table 3b). In GS, h^2_{SNP} was not significant for GHQ (h^2_{SNP} = 0.071, p = 0.165; N = 4,919). However, in an ad hoc estimation from the baseline sample of 6,751 unrelated GS participants (details in **Appendix D**: Supplementary Table 3b) we detected a significant h^2_{SNP} for GHQ (h^2_{SNP} = 0.135; $p < 5.15 \times 10^{-3}$), suggesting that the power to estimate h^2_{SNP} in GS may be limited by sample size. Estimates were not significant for neither TSLE (h^2_{SNP} = 0.061, p = 0.189; **Appendix D**: Supplementary Table 3b) nor ISLE (h^2_{SNP} = 0.000, p = 0.5), but h^2_{SNP} was significant for DSLE (h^2_{SNP} = 0.131, p = 0.029), supporting a potential genetic mediation and gene-environment correlation.

5.4.3 GWAS of depressive symptoms

No genome-wide significant SNPs were detected by GWAS in either cohort. Top findings ($p < 1 \times 10^{-5}$) are summarized in **Appendix D**: Supplementary Table 4. Manhattan and QQ plots are shown in **Appendix D**: Supplementary Figures 1-4. There was no evidence of genomic inflation (all $\lambda_{1000} < 1.01$).

5.4.4 Post-GWAS analyses

Gene-based test identified six genes associated with PHQ using UKB-GWAS statistics at genome-wide significance (Bonferroni-corrected $p = 2.77 \times 10^{-6}$; *DCC*, $p = 7.53 \times 10^{-8}$; *ACSS3*, $p = 6.51 \times 10^{-7}$; *DRD2*, $p = 6.55 \times 10^{-7}$; *STAG1*, $p = 1.63 \times 10\text{-}^{6}$; *FOXP2*, $p = 2.09 \times 10^{-6}$; *KYNU*, $p = 2.24 \times 10^{-6}$; **Appendix D**: Supplementary Figure 8). Prioritized genes based on position, eQTL and chromatin interaction mapping are detailed in **Appendix D**: Supplementary Table 5. No genes were detected in GS-GWAS gene-based test (**Appendix D**: Supplementary Figures 9). No tissue enrichment was detected from GWAS in either cohort. Significant gene-sets and GWAS catalog associations for UKB-GWAS are reported in **Appendix D**: Supplementary Table 6. These included the *biological process*: positive regulation of long term synaptic potentiation, and *GWAS catalog associations*: brain structure, schizophrenia, response to amphetamines, age-related cataracts (age at onset), fibrinogen, acne (severe), fibrinogen levels, and educational attainment; all adjusted-$p < 0.01$. There was no significant gene-set enrichment from GS-GWAS.

5.4.5 GWEIS of depressive symptoms

Manhattan and QQ plots are shown in **Appendix D**: Supplementary Figures 1-4. There was no evidence of GWEIS inflation for either UKB or GS (all $\lambda_{1000} < 1.01$). No genome-wide significant GWEIS associations were detected for SLE in UKB. GS-GWEIS using TSLE identified a significant GxE effect ($p < 2.97 \times 10^{-8}$) at an intragenic SNP on chromosome 11 (rs12789145, $p = 4.95 \times 10^{-9}$, $\beta = 0.06$, closest gene: *PIWIL4*; **Appendix D**: Supplementary Figure 5), and using DSLE at an intronic SNP in *ZCCHC2* on chromosome 18 (rs17070072, $p = 1.46 \times 10^{-8}$, $\beta = -0.08$; **Appendix D**: Supplementary Figure 6). In their corresponding joint effect tests both rs12789145 ($p = 2.77 \times 10^{-8}$) and rs17070072 $p = 1.96 \times 10^{-8}$) were significant. GWEIS for joint effect using DSLE identified two further significant SNPs on chromosome 9 (rs12000047, $p = 2.00 \times 10^{-8}$, $\beta = -0.23$; rs12005200, $p = 2.09 \times 10^{-8}$, $\beta = -0.23$, LD $r^2 > 0.8$, closest gene: *CYLC2*; **Appendix D**: Supplementary Figure 7). None of these associations replicated in UKB ($p > 0.05$), although the

effect direction was consistent between cohorts for the SNP close to *PIWL1* and SNPs at *CYLC2*. No SNP achieved genome-wide significant association in GS-GWEIS using ISLE as exposure. Top GWEIS results ($p < 1 \times 10^{-5}$) are summarized in **Appendix D**: Supplementary Tables 7-10.

5.4.6 Post-GWEIS analyses: gene-based tests

All results are shown in **Appendix D**: Supplementary Figures 10-17. Two genes were associated with PHQ using the joint effect from UKB-GWEIS (*ACSS3* $p = 1.61 \times 10^{-6}$; *PHF2*, $p = 2.28 \times 10^{-6}$; **Appendix D**: Supplementary Figure 11). *ACSS3* was previously identified using the additive main effects, whereas *PHF2* was only significantly associated using the joint effects. Gene-based tests identified *MTNR1B* as significantly associated with GHQ on GS-GWEIS using DSLE in both GxE ($p = 1.53 \times 10^{-6}$) and joint effects ($p = 2.38 \times 10^{-6}$; **Appendix D**: Supplementary Figures 14-15).

5.4.7 Post-GWEIS analyses: tissue enrichment

We prioritized genes based on position, eQTL and chromatin interaction mapping in brain tissues and regions. In UKB, prioritized genes with GxE effect were enriched for up-regulated differentially expressed genes from adrenal gland (adjusted-$p = 3.58 \times 10^{-2}$). Using joint effects, prioritized genes were enriched on up-regulated differentially expressed genes from artery tibial (adjusted-$p = 4.34 \times 10^{-2}$). In GS, prioritized genes were enriched: in up-regulated differentially expressed genes from artery coronary (adjusted-$p = 4.55 \times 10^{-2}$) using GxE effects with DSLE; in down-regulated differentially expressed genes from artery aorta tissue (adjusted-$p = 4.71 \times 10^{-2}$) using GxE effects with ISLE; in up-regulated differentially expressed genes from artery coronary (adjusted-$p = 5.97 \times 10^{-3}$, adjusted-$p = 9.57 \times 10^{-3}$) and artery tibial (adjusted-$p = 1.05 \times 10^{-2}$, adjusted-$p = 1.55 \times 10^{-2}$) tissues using joint effects with both TSLE and DSLE; and in down-regulated differentially expressed genes from lung tissue (adjusted-$p = 3.98 \times 10^{-2}$) and in up- and down-regulated differentially expressed genes from the spleen (adjusted-$p = 4.71 \times 10^{-2}$) using joint effects with ISLE. There was no enrichment using GxE effect with TSLE.

5.4.8 Post-GWEIS analyses: gene-sets enrichment

Significant gene-sets and GWAS catalog hits from GWEIS are detailed in **Appendix D**: Supplementary Tables 11-14, including for UKB *Biocarta*: GPCR pathway; *Reactome*: opioid signalling, neurotransmitter receptor binding and downstream transmission in the postsynaptic cell, transmission across chemical synapses, gastrin CREB signalling pathway via PKC and MAPK; *GWAS catalog*: post bronchodilator FEV1/FVC ratio, migraine and body mass index. In GS, enrichment was seen using TSLE and DLSE for *GWAS catalog*: age-related macular degeneration, myopia, urate levels and Heschl's gyrus morphology; and using ISLE for *biological process*: regulation of transporter activity. All adjusted-$p < 0.01$.

5.4.9 Cross-cohort prediction

In GS, PRS_D weighted by UKB-GWAS of PHQ significantly explained 0.56% of GHQ variance (*Empirical-p* < 1.10^{-4}), similar to PRS_{MDD} weighted by PGC MDD-GWAS (R^2 = 0.78%, *Empirical-p* < 1.10^{-4}). PRS_{GxE} weighted by UKB-GWEIS GxE effects explained 0.15% of GHQ variance (*Empirical-p* = 0.03, **Appendix D**: Supplementary Table 15). PRS_{GxE} fitted jointly with PRS_{MDD} significantly improved prediction of GHQ (R^2 = 0.93%, model $p = 6.12 \times 10^{-11}$; predictive improvement of 19%, *LRT-p* = 5.91×10^{-3}) compared to PRS_{MDD} alone. Similar to PRS_{GxE} with PRS_D (R^2 = 0.69%, model $p = 2.72 \times 10^{-8}$; predictive improvement of 23%, *LRT-p* = 0.01). PRS_{Joint} weighted by UKB-GWEIS also predicted GHQ (R^2 = 0.58%, *Empirical-p* < 1.10^{-4}), although the variance explained was significantly reduced compared to the model fitting PRS_{GxE} and PRS_D together (*LRT-p* = 4.69×10^{-7}), suggesting that additive and GxE effects should be modelled independently for polygenic approaches (**Figure 5.2a**).

In UKB (**Figure 5.2b**), both PRS_D weighted by GS-GWAS of GHQ and PRS_{MDD} significantly explained 0.04% and 0.45% of PHQ variance, respectively (both *Empirical-p* < 1.10^{-4}; **Appendix D**: Supplementary Table 15). PRS_{GxE} derived from GS-GWEIS GxE effect did not significantly predicted PHQ (TSLE-PRS_{GxE} *Empirical-p* = 0.382; DSLE-PRS_{GxE} *Empirical-*

p = 0.642; ISLE-PRS_{GxE} *Empirical-p* = 0.748). Predictive improvements by PRS_{GxE} effect fitted jointly with PRS_{MDD} or PRS_D were not significant (all *LRT-p* > 0.08). PRS_{Joint} significantly predicted PHQ (TSLE-PRS_{Joint}: R^2 = 0.032%, *Empirical-p* < 1.10^{-4}; DSLE-PRS_{Joint}: R^2 = 0.012%, *Empirical-p* = 4.3 x 10^{-3}; ISLE-PRS_{Joint}: R^2 = 0.032%, *Empirical-p* < 1.10^{-4}), although the variances explained were significantly reduced compared to the models fitting PRS_{GxE} and PRS_D together (all *LRT-p* < 1.48 x 10^{-3}).

5.4.10 Prediction of stress-related traits

Prediction of stress-related traits in independent samples using PRS_D, PRS_{GxE} and PRS_{Joint} are summarized in **Figure 5.3a** and **Appendix D**: Supplementary Table 16. Significant trait prediction after FDR adjustment (*Empirical-p* < 6.16 x 10^{-3}, FDR-adjusted *Empirical-p* < 0.05) using both UKB and GS PRS_D was seen for: depression status, neuroticism and schizotypal personality. PRS_{GxE} weighted by GS-GWEIS GxE effect using ISLE significantly predicted depression status mapping by proxy (*Empirical-p* = 7.00 x 10^{-4}, *FDR-adjusted Empirical-p* = 9.54 x 10^{-3}).

Nominally significant predictive improvements (*LRT-p* < 0.05) of fitting PRS_{GxE} over the PRS_D effect alone using summary statistics generated from both UKB and GS were detected for schizotypal personality, heart diseases and COPD by proxy (**Figure 5.3b**). PRS_{GxE} weighted by GS-GWEIS GxE effect using ISLE significantly improved prediction over PRS_D effect alone of depression status mapping by proxy after FDR adjustment (*LRT-p* = 1.96 x 10^{-4}, *FDR-adjusted LRT-p* = 2.35 x 10^{-2}).

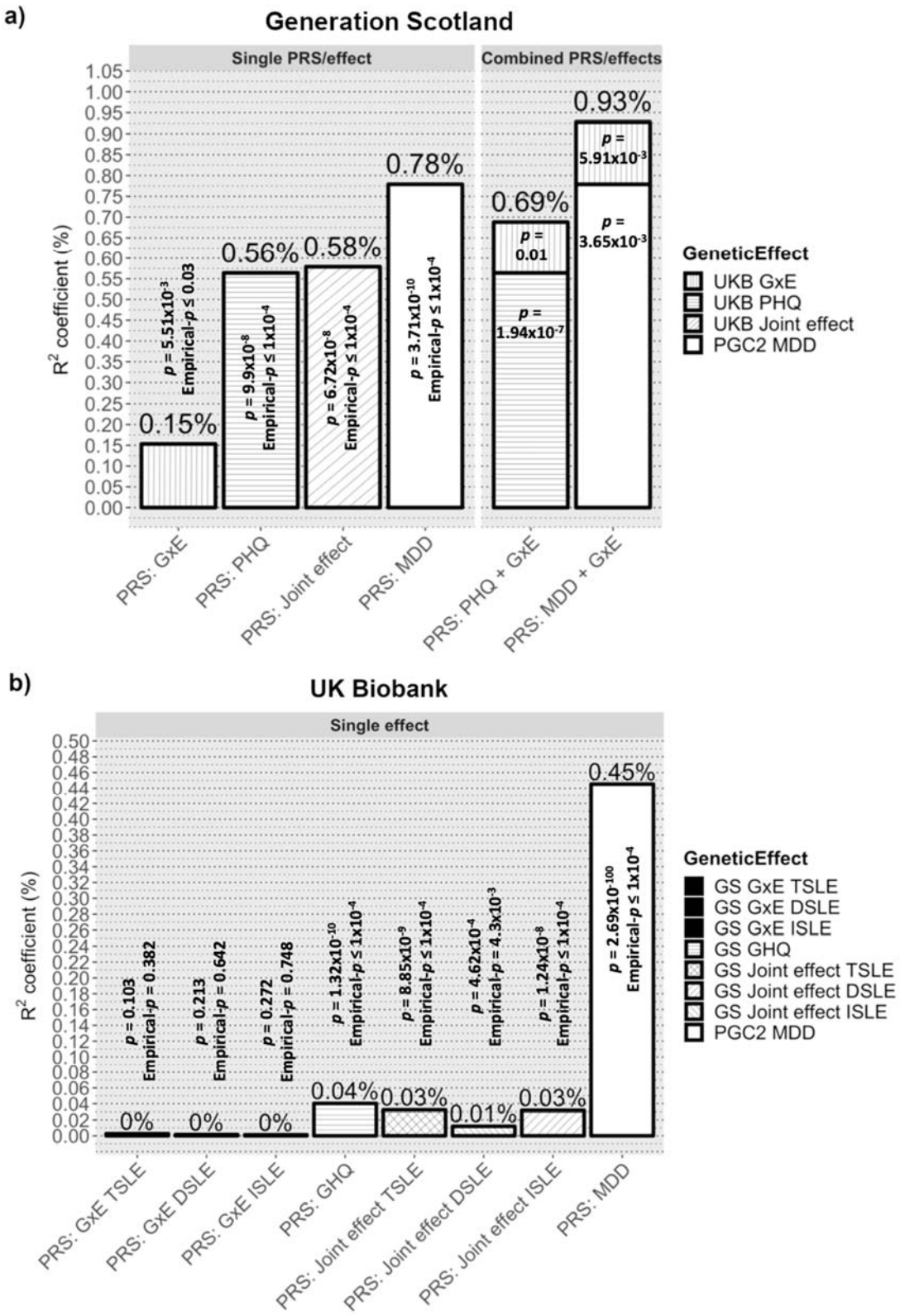

a)
Generation Scotland
Single PRS/effect
Combined PRS/effects
R2 coefficient (%)
0.15%
p = 5.51x10-3
Empirical-p ≤ 0.03
0.56%
p = 9.9x10-8
Empirical-p ≤ 1x10-4
0.58%
p = 6.72x10-8
Empirical-p ≤ 1x10-4
0.78%
p = 3.71x10-10
Empirical-p ≤ 1x10-4
0.69%
p = 0.01
p = 1.94x10-7
0.93%
p = 5.91x10-3
p = 3.65x10-3
PRS: GxE
PRS: PHQ
PRS: Joint effect
PRS: MDD
PRS: PHQ + GxE
PRS: MDD + GxE
GeneticEffect
UKB GxE
UKB PHQ
UKB Joint effect
PGC2 MDD
b)
UK Biobank
Single effect
R2 coefficient (%)
0%
p = 0.103
Empirical-p = 0.382
0%
p = 0.213
Empirical-p = 0.642
0%
p = 0.272
Empirical-p = 0.748
0.04%
p = 1.32x10-10
Empirical-p ≤ 1x10-4
0.03%
p = 8.85x10-9
Empirical-p ≤ 1x10-4
0.01%
p = 4.62x10-4
Empirical-p = 4.3x10-3
0.03%
p = 1.24x10-8
Empirical-p ≤ 1x10-4
0.45%
p = 2.69x10-100
Empirical-p ≤ 1x10-4
PRS: GxE TSLE
PRS: GxE DSLE
PRS: GxE ISLE
PRS: GHQ
PRS: Joint effect TSLE
PRS: Joint effect DSLE
PRS: Joint effect ISLE
PRS: MDD
GeneticEffect
GS GxE TSLE
GS GxE DSLE
GS GxE ISLE
GS GHQ
GS Joint effect TSLE
GS Joint effect DSLE
GS Joint effect ISLE
PGC2 MDD

Figure 5.2 Prediction of depression scores by PRS_{GxE}, PRS_{D}, PRS_{MDD} & PRS_{Joint}. *Variance of depression score explained by PRS_{GxE} PRS_{D}, PRS_{MDD} and PRS_{Joint} as single effect; and combining both PRS_{D} and PRS_{MDD} with PRS_{GxE} in single models. Prediction was conducted using **2a)** Generation Scotland and **2b)** UK Biobank as target sample. PRS_{GxE} were weighted by cross sample GWEIS using GxE effect. PRS_{D} were weighted by cross sample GWAS of depressive symptoms effect. PRS_{MDD} was weighted by PGC MDD-GWAS summary statistics. PRS_{Joint} were weighted by cross sample GWEIS using joint effect. A nominally significant gain in variance explained of GHQ of about 23% was seen in Generation Scotland when PRS_{GxE} was incorporated into a multiple regression model along with PRS_{D}; and of about 19% when PRS_{GxE} was incorporated into a multiple regression model along with PRS_{MDD}. Such gain was not seen in UK Biobank, but it must be noted that both PRS_{D} and PRS_{MDD} also explains much less variance of PHQ in UK Biobank than of GHQ in Generation Scotland. To note a noticeably reduction of variance explained by PRS_{Joint} compared to combined PRS/effects.*

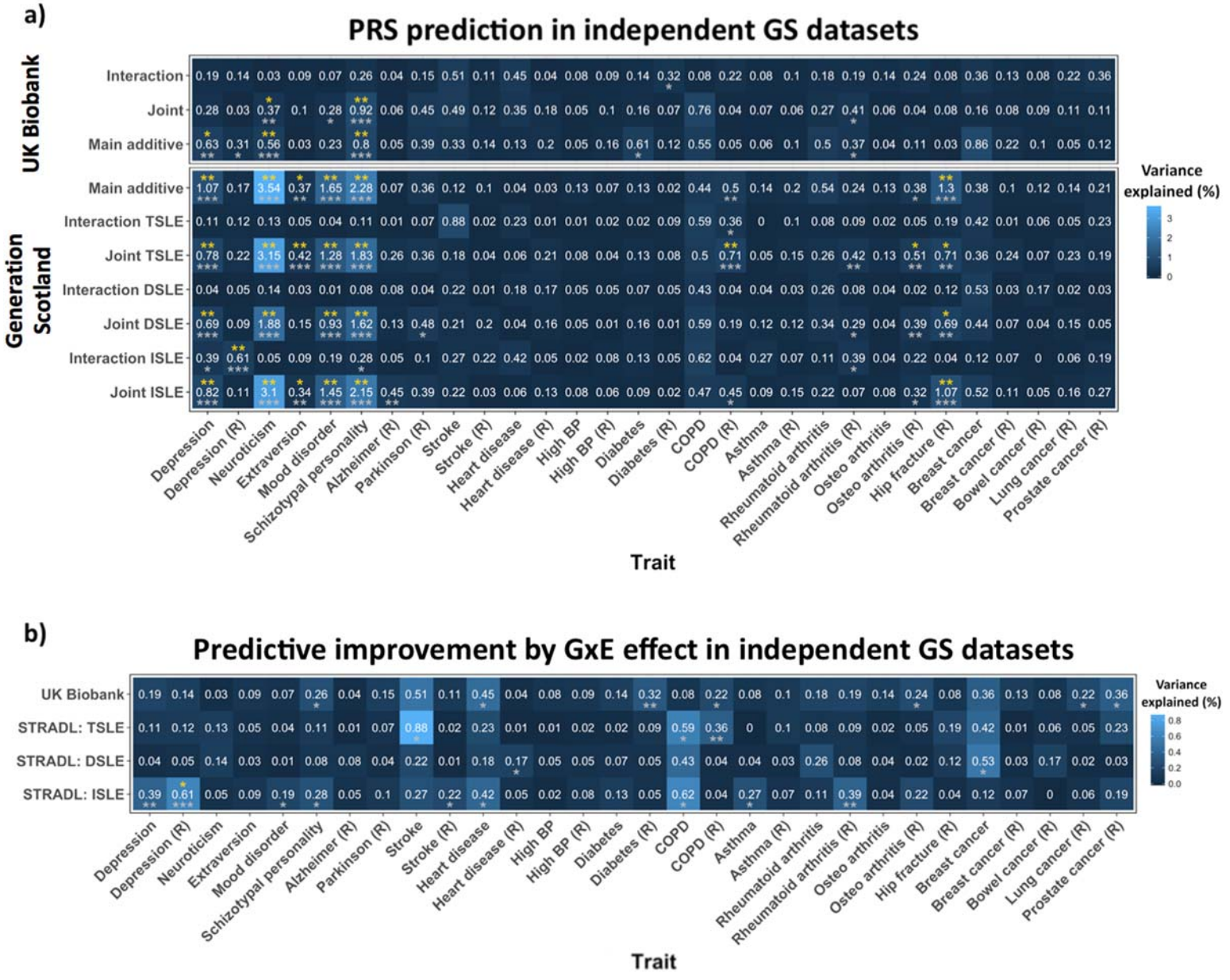

a)
PRS prediction in independent GS datasets
UK Biobank
Interaction
Joint
Main additive
Generation Scotland
Main additive
Interaction TSLE
Joint TSLE
Interaction DSLE
Joint DSLE
Interaction ISLE
Joint ISLE
Depression
Depression (R)
Neuroticism
Extraversion
Mood disorder
Schizotypal personality
Alzheimer (R)
Parkinson (R)
Stroke
Stroke (R)
Heart disease
Heart disease (R)
High BP
High BP (R)
Diabetes
Diabetes (R)
COPD
COPD (R)
Asthma
Asthma (R)
Rheumatoid arthritis
Rheumatoid arthritis (R)
Osteo arthritis
Osteo arthritis (R)
Hip fracture (R)
Breast cancer
Breast cancer (R)
Bowel cancer (R)
Lung cancer (R)
Prostate cancer (R)
Trait
Variance explained (%)
b)
Predictive improvement by GxE effect in independent GS datasets
UK Biobank
STRADL: TSLE
STRADL: DSLE
STRADL: ISLE
Trait
Variance explained (%)

Figure 5.3 a) PRS prediction in independent GS datasets. *Heatmap illustrating PRS prediction of a wide range of traits from GS listed in the x-axis (**Table 5.1**). (R) refers to traits using mapping by proxy approach (i.e. where first-degree relatives of individuals with the disease are considered proxy cases and included into the group of cases). Y-axis shows the discovery sample and the effect used to weight PRS. Numbers in cells indicate the % of variance explained, also represented by colour scale. Significance is represented by "*" according to the following significance codes: p-values **** < 1 x 10^{-4} < *** < 0.001 < ** < 0.01 * < 0.05; in grey Empirical-p-values after permutation (10,000 times) and in yellow FDR-adjusted Empirical-p-values.* **b) Predictive improvement by GxE effect in independent GS datasets.** *Heatmanp illustrating the predictive improvement as a result of incorporating PRS_{GxE} into a multiple model along with PRS_D and covariates (full model), over a model fitting PRS_D alone with covariates (null model); predicting a wide range of traits from GS listed in the x-axis (**Table 5.1**). Covariates: age, sex and 20 PCs. (R) refers to traits using mapping by proxy approach (i.e. where first-degree relatives of individuals with the disease are consider proxy cases and included into the group of cases). PRS_{GxE} are weighted by GWEIS using GxE effects. PRS_D were weighted by the GWAS of depressive symptoms additive main effects. The Y-axis shows the discovery sample used to weight PRS. Numbers in cells indicate the % of variance explained by the PRS_{GxE}, also represented by colour scale. Notice that those correspond to the PRS_{GxE} predictions in **Figure 5.3a** when PRS_{GxE} are weighted by GxE effects. Significance was tested by Likelihood ratio tests (LRT): full model including PRS_D + PRS_{GxE} vs. null model with PRS_D alone (covariates adjusted). Significance is represented by "*" according to the following significance codes: p-values *** < 0.001 < ** < 0.01 * < 0.05; in grey LRT-p-values and in yellow FDR-adjusted LRT-p-values.*

5.5 Discussion

This study performs GWAS and incorporates data on recent adult stressful life events (SLE) into GWEIS of depressive symptoms, identifies new loci and candidate genes for the modulation of genetic response to SLE; and provides insights to help disentangle the underlying aetiological mechanisms increasing genetic liability through SLE to both depressive symptoms and stress-related traits.

SNP-heritability of depressive symptoms (h^2_{SNP} = 9-13%), were slightly higher than estimates from African American populations[113], and over a third larger than estimates in MDD from European samples[182]. h^2_{SNP} for PHQ in UKB (9.0%) remained significant after adjusting for SLE (7.9%). Thus, although some genetic contributions may be partially shared between depressive symptoms and reporting of SLE, there is still a relatively large genetic contribution unique to depressive symptoms. Significant h^2_{SNP} of DSLE in GS (13%) and $TSLE_{UKB}$ in UKB (4%), which is mainly composed of dependent SLE items, were detected similar to previous studies (8% and 29%)[112,113]. Conversely, there was no evidence for heritability of independent SLE. A significant bivariate genetic correlation between depressive symptoms and SLE (rG = 0.72) was detected in UKB after adjusting for covariates, suggesting that there are shared common variants underlying self-reported depressive symptoms and SLE. This bivariate genetic correlation was smaller than that estimated from African American populations (rG = 0.97; p= 0.04; N = 7,179)[113]. Genetic correlations between SLE measures and GHQ were not significant in GS (N = 4,919; rG = 1; all p > 0.056), perhaps due to a lack of power in this smaller sample.

Post-GWAS gene-based tests detected six genes significantly associated with PHQ (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*). Previous studies have implicated these genes in liability to depression (see **Appendix D**: Supplementary Table 17), and three of them are genome-wide significant

in gene-based tests from the latest meta-analysis of major depression that includes UKB (*DCC*, $p = 2.57 \times 10^{-14}$; *DRD2*, $p = 5.35 \times 10^{-14}$; and *KYNU*, $p = 2.38 \times 10^{-6}$; N = 807,553)[151]. This supports the implementation of quantitative measures such as PHQ to detect genes underlying lifetime depression status[449]. For example, significant gene ontology analysis of the UKB-GWAS identified enrichment for positive regulation of long-term synaptic potentiation, and for previous GWAS findings of brain structure[450], schizophrenia[451] and response to amphetamines[452].

The key element of this study was to conduct GWEIS of depressive symptoms and recent SLE. We identified two loci with significant GxE effect in GS. However, none of these associations replicated in UKB ($p > 0.05$). The strongest association was using TSLE at 53kb down-stream of *PIWIL4* (rs12789145). PIWIL4 is brain-expressed and involved in chromatin-modification[453], suggesting it may moderate the effects of stress on depression. It encodes HIWI2, a protein thought to regulate OTX2, which is critical for the development of forebrain and for coordinating critical periods of plasticity disrupting the integration of cortical circuits[454,455]. Indeed, an intronic SNP in *PIWIL4* was identified as the strongest GxE signal in ADHD using mother's warmth as environmental exposure[456]. The other significant GxE identified in our study was in *ZCCHC2* using DSLE. This zinc finger protein is expressed in blood CD4+ T-cells and is down regulated in individuals with MDD[457] and in those resistant to treatment with citalopram[458]. No GxE effect was seen using ISLE as exposure.

No significant locus or gene with GxE effect was detected in UKB-GWEIS. Nevertheless, joint effects (combined additive main and GxE genetic effects) identified two genes significantly associated with PHQ (*ACSS3* and *PHF2*; see **Appendix D**: Supplementary Table 17). *PHF2* was recently detected as genome-wide significant at the latest meta-analysis of depression[151]. Notably, *PHF2* paralogs have already been link with MDD through stress-response in three other studies[343,344,459]. Joint effects in GS also detected an additional significant association upstream *CYLC2*, a gene nominally associated ($p < 1$

x 10^{-5}) with obsessive-compulsive disorder and Tourette's Syndrome[460]. Gene-based test from GS-GWEIS identified a significant association with *MTNR1B,* a melatonin receptor gene, using DSLE (both GxE and joint effect; **Appendix D**: Supplementary Table 17). Prioritized genes using GxE effects were enriched in differentially expressed genes from several tissues including the adrenal gland, which releases cortisol into the blood-stream in response to stress, thus playing a key role in the stress-response system, reinforcing a potential role of GxE in stress-related conditions.

The different instruments and sample sizes available make it hard to compare results between cohorts. Whereas GS contains deeper phenotyping measurements of stress and depressive symptoms than UKB, the sample size is much smaller, which may be reflected in the statistical power required to detect reliable GxE effects. Furthermore, the presence and size of GxE are dependent on their parameterization (i.e. the measurement, scale and distribution of the instruments used to test such interaction)[425]. Thus, GxE may be incomparable across GWEIS due to differences in both phenotype assessment and stressors tested. Although our results suggest that both depressive symptom measures are correlated with lifetime depression status, different influences on depressive symptoms from the SLE covered across studies may contribute to lack of stronger replication. Instruments in GS cover a wider range of SLE and are more likely to capture changes in depressive symptoms as consequence of their short-term effects. Conversely, UKB could capture more marked long-term effects, as SLE were captured over 2 years compared to 6 months in GS. New mental health questionnaires covering a wide range of psychiatric symptoms and SLE in the last release of UKB data provides the opportunity to create more similar measures to GS in the near future. Further replication in independent studies with equivalent instruments is required to validate our GWEIS findings. Despite these limitations and a lack of overlap in the individual genes prioritised from the two GWEIS, replication was seen in the predictive improvement of using PRS_{GxE} derived from the GWEIS GxE effects to predict stress-related phenotypes.

The third aim of this study was to test whether GxE effect could improve predictive genetic models, and thus help to explain deviation from additive models and missing heritability of MDD[82]. Multiple regression models suggested that inclusion of PRS_{GxE} weighted by GxE effects could improve prediction of an individual's depressive symptoms over use of PRS_{MDD} or PRS_{D} weighted by additive effects alone. In GS, we detected a predictive gain of 19% over PRS_{MDD} weighted by PGC MDD-GWAS, and a gain of 23% over PRS_{D} weighted by UKB-GWAS (**Figure 5.2a**). However, these findings did not surpass stringent Bonferroni-correction and could not be validated in UKB. This may reflect in the poor predictive power of the PRS generated from the much smaller GS discovery sample. The results show a noticeably reduced prediction using PRS_{Joint} weighted by joint effects, which suggests that the genetic architecture of stress-response is at least partially independent and differs from genetic additive main effects. Therefore, our results from multiple regression models suggest that for polygenic approaches main and GxE effects should be modelled independently.

SLE effects are not limited to mental illness[444]. Our final aim was to investigate shared aetiology between GxE for depressive symptoms and stress-related traits. Despite the differences between the respondents and non-respondents (**Table 5.1** legend), a significant improvement was seen predicting depressive status mapping by proxy cases using GxE effect from GS-GWEIS with independent SLE (*FDR-adjusted LRT-p* = 0.013), but not with dependent SLE. GxE effects using statistics generated from both discovery samples, despite the differences in measures, nominally improved the phenotypic prediction of schizotypal personality, heart disease and the proxy of COPD (*LRT-p* < 0.05). Other studies have found evidence supporting a link between stress and depression in these phenotypes that support our results (see **Appendix D.3** for extended review) and suggest, for instance, potential pleiotropy between schizotypal personality and stress-response. Our findings point to a potential genetic component underlying a stress-response-depression-comorbidities link due, at least in part, to shared stress-response mechanisms. A relationship between SLE, depression and

coping strategies such as smoking suggests that perhaps, genetic stress-response may modulate adaptive behaviours such as smoking, fatty diet intake, alcohol consumption and substance abuse. This is discussed further in the **Appendix D.3**.

In this study, evidence for SNPs with significant GxE effects came primarily from the analyses of dependent SLE and not from independent SLE. This supports a genetic effect on probability of exposure to, or reporting of SLE, endorsing a gene-environment correlation. Chronic stress may influence cognition, decision-making and behaviour eventually leading to higher risk-taking[461]. These conditions may also increase sensitivity to stress amongst vulnerable individuals, including those with depression, who also have a higher propensity to report SLE, particularly dependent SLE[98]. A potential reporting bias in dependent SLE may be mediated as well by heritable behavioural, anxiety or psychological traits such as risk-taking[112,462]. Furthermore, individuals vulnerable to MDD may expose themselves into environments of higher risk and stress[103]. This complex interplay, reflected in the form of a gene-environment correlation effect, would hinder the interpretation of GxE effects from GWEIS as pure interactions. A mediation of associations between SLE and depressive symptoms through genetically driven sensitivity to stress, personality or behavioural traits would support the possibility of subtle genotype-by-genotype (GxG) interactions, or genotype-by-genotype-by-environment (GxGxE) interactions contributing to depression[463,464]. In contrast, PRS prediction of the stress-related traits: schizotypal personality, heart disease and COPD, was primarily from derived weights using independent SLE, suggesting that a common set of variants moderate the effects of SLE across stress-related traits and that larger sample sizes will be required to detect the individual SNPs contributing to this. Thus, our finding supports the inclusion of environmental information into GWAS to enhance the detection of relevant genes. Results of studying dependent and independent SLE support a contribution of genetically mediated exposure to and/or reporting of SLE, perhaps through sensitivity to stress as mediator.

This study emphasises the relevance of GxE in depression and human health in general and provides the basis for future lines of research.

Chapter 6 A new test of the diathesis-stress framework for depression: contributions to liability from genetics underlying sensitivity and response to psychosocial stress

I have generated summary data on two distinct SNP effects related to genetic response to environmental stress: the effect of a proxy for MDD-dependent sensitivity to stress derived in **chapter 2**, and the GxE effect to reported SLE on depression score derived in **chapter 5**. In **chapter 4**, I implemented a *diathesis-stress* framework in order to investigate GxE effects in depression. However, the SNP effects used to weight the PRS, included in the PRSxSLE term, were derived from PGC GWAS of MDD and, therefore, based in the main additive effects of MDD. In this chapter, I incorporate the data I generated in **chapter 2** and **chapter 5** in order to test the *diathesis-stress* model implemented in **chapter 4** using alternative weightings for the PRS. I implement new tests of the *diathesis-stress* model for depression using weightings for stress-sensitivity and stress-response in order to test for the presence of GxE effect with reported SLE, and compare these results to those in **chapter 4**.

The following chapter has been formatted for publication to *PLOS Genetics*, which explains the use of "we" along the chapter and the inclusion of an author summary. I confirm that the work presented is my own work under guidance from my supervisor Dr. Pippa Thomson. I performed all the analyses myself.

Expected publication:

Arnau-Soler, A. *et al.* A new test of the diathesis-stress framework for depression: contributions to liability from genetics underlying sensitivity and response to psychosocial stress (2019).

6.1 Abstract

Gene-by-environment interactions (GxE) for risk of depression have been evaluated under *diathesis-stress* models assuming that most genetic variants with additive effects also contribute to GxE. An alternative hypothesis is that GxE effects may be displayed by variants with negligible additive effects.

We implement polygenic risk scores (PRS) weighted by either a Genome-Wide Interaction Study of a proxy for stress-sensitivity (PRS_{SS}) or a Genome-Wide by Environment Interaction Study for depressive symptoms (PRS_{GxE}) to predict depression, both independently and combined with PRS for major depressive disorder (PRS_{MDD}) and reported stressful life events (SLE), in a sample of 4,919 unrelated individuals. $SLExPRS_{MDD}$, $SLExPRS_{SS}$, and $SLExPRS_{GxE}$ interactions are used to test *diathesis-stress* theory for depression, and differences between women and men examined.

Both PRS_{SS} and PRS_{GxE} predict liability to depressive symptoms not accounted for by PRS_{MDD}. A nominally significant $SLExPRS_{MDD}$ interaction was detected in the full sample, but not with $SLExPRS_{SS}$ or $SLExPRS_{GxE}$. In comparison, a previous subset of the discovery sample for MDD did not detect a significant $SLExPRS_{MDD}$ effect, suggesting that with large enough discovery samples more robust GxE effects could be achieved. We detected a significant $SLExPRS_{SS}$ interaction in men explaining 0.42% of the variance in depression (p = 2.1 x 10^{-3}) in addition to 0.50% explained by PRS_{MDD}. The effect of $SLExPRS_{SS}$ was significantly different across sexes (p = 1.00 x 10^{-3}). Categorizing the number of SLE reported in none, low (1 or 2) and high (3 or more), PRS_{SS} increased the risk of depression in men who reported high numbers of SLE (p = 1.84 x 10^{-4}), following similar patterns in the full sample (p = 7.37 x 10^{-3}) and in women (p = 0.03).

Our study investigates the potential of two genome-wide alternatives to standard GWAS, for the detection of common variants with effects related to

stress response significantly contributing to depression, and for using its estimated effects as new weights to test the *diathesis-stress* theory.

6.2 Author summary

Stress has implications for many human conditions, including depression, a complex psychiatric disorder that requires from both genetic and environmental factors to manifest. Standard genome-wide association studies have identified common genetic variants displaying a direct association with depression. However, such variants may be distinct from those responding to the effects of stress and undetectable when the environment is not taken in to account. Here we show that common variants with stress-related effects add an additional risk of suffering depressive symptoms. Our findings show that people, particularly men, are at higher risk of depression due to genetic sensitivity to stress when they report high numbers of recent stressful life events. The results suggest an important contribution to psychiatric genetic research of alternative genome-wide approaches as discovery sample sizes increase. Identifying the presence of genetic pathways with modulatory effects to psychological stress could lead to better treatments and prevention strategies not only in depression, but also in other stress-related conditions.

6.3 Introduction

Major depressive disorder (MDD) develops from the complex interplay of many genetic factors forming an individual's polygenic vulnerability and the environment[119,226,465]. Twin studies estimate that for MDD, genetic factors account for about 40% of the phenotypic variance[119], the contribution of common genetic variants in liability to MDD is approximately 10%[150], and genome-wide association studies (GWAS) have already identified genetic risk variants robustly associated with MDD[149,150,152,153,184]. However, results from twin studies estimate that environmental influences account for up to 63% of the liability to MDD[119], of which stressful life events (SLE) are among the environmental factors most consistently associated[84,90,94,402].

The *diathesis-stress* theory conceptualizes the development of psychopathologies as the combined effect of genetic and environmental risk factors[228]. It suggests that individuals have a latent inherent vulnerability or predisposition to depression, a so-called *diathesis,* that is variable across individuals[225]. However, it assumes that this *diathesis* is not enough and some environmental trigger, such as SLE, is required for the development of symptoms[225]. According to the theory, people with a higher *diathesis* for MDD will require lower amounts of stress to develop symptoms, and vice versa. However, these additive effects may not be enough and some additional effect may be required to exceed the threshold at which the clinical disorders manifest. Thus, the *diathesis-stress* theory proposes the presence of a genotype-by-environment interaction (GxE) effect between an individual's vulnerability and psychological stress that plays a key role in the onset and development of MDD. However, the nature of the GxE effect underpinning the development of depressive symptoms remains elusive.

Polygenic risk scores (PRS) are quantitative metrics of an individual's overall genetic vulnerability aggregating the additive effects of many common genetic risk variants[466]. Therefore, PRS can be conceptualized as indicators

of the *diathesis* (i.e. genetic vulnerability or predisposition to disease). This allows the *diathesis-stress* theory to be empirically tested in a model that takes into account the *diathesis* effect (operationalized as PRS), the *stress* effect of SLE (operationalized as a self-reported measure of SLE), and the *diathesis-stress* interaction effect (operationalized as a GxE interaction product of $SLExPRS$). Thus providing an empirical method to seek evidence of GxE effects in liability to MDD. A PRSxE approach has successfully identified GxE effects in other traits such as obesity with physical activity[467] or with sugar sweetened beverages[468]. Despite the reduction in statistical power caused by using a PRS that may include SNPs without a truly modulatory effect, the gain in explanatory power of PRS compared to single SNPs remains considerable[274].

Following the expansion and success of GWAS and polygenic approaches, GxE studies have empirically tested this *diathesis-stress* model for depression with adult SLE[86,169,276,431], reporting the first evidence of a genome-wide GxE effect[169,431]. However, a plausible explanation for the weakly significant GxE effects identified may be the rationale behind the construction of the *diathesis-stress* model in these studies. These studies generated PRS using the additive effect at each SNP for MDD from Psychiatric Genomics Consortium genome-wide association studies (PRS_{MDD}) to weight both the additive and GxE terms. However, some genetic factors contributing to GxE effects in depression may not display additive main contributions to MDD in the absence of stress. Indeed, there is evidence that polymorphisms in genes such as *BDNF* and *HTR1A*, may only influence the development of MDD through a direct effect on vulnerability to the effects of stress and/or environmental adversity[464,469-473]. In accordance with this, the expression of *BDNF* in the hippocampus of rats is dependent on the amount of stress[474] and important for long-term synapse potentiation[475]. Nevertheless, *BDNF* was not identified in the latest GWAS of MDD involving 807,553 individuals[151]. Hence, the effect of genetic variants with stress-related effects on risk of depression may remain hidden in current GWAS that do not take into account environmental influences. It also suggests that

the SNPs involved in the additive main effects and the GxE effects may correspond to different biological pathways. An alternative model could be built using distinct PRS to operationalize the main *diathesis* effect and the interaction effect with SLE for depression. In this study we examine two possible alternate weightings for the GxE term derived from studies of stress-sensitivity and genome-wide by environment interaction studies (GWEIS)[378].

In a previous study, we generated a proxy for individual's sensitivity to negative outcomes by modelling the interaction between SNP allele and MDD status on neuroticism score through a genome-wide interaction study (GWIS) [378]. Individuals with MDD score higher in tests of neuroticism[213,214], an heritable personality trait[476] thought to mediate or moderate the effects of adverse experiences[79,222]. This GWIS identified common genetic variants that showed MDD-dependent effects on neuroticism and implicated glucocorticoid receptor function as a mediator of sensitivity to stress. PRS weighted by the genetic contributions to such proxy for stress-sensitivity (PRS_{SS}) significantly predicted the liability to MDD not attributable to additive main effects of either MDD or neuroticism[378].

External weights from GWEIS may also increase the power to detect GxE[477]. GWEIS evaluate the association between a trait of interest and the GxE effect of a SNP allele and an environmental risk factor at the genome-wide scale. To date, only a very few GWEIS of depression and SLE are available[113,267-269,427]. All of which use relatively small sample sizes (N = 320 - 99,057), despite this, all report at least one significant locus. PRS weighted by GxE effects (PRS_{GxE}) using the results of GWEIS significantly improve the prediction of depressive symptoms[427].

In this study, we aim to test whether the non-additive genetic contributions are relevant in predicting the risk of depression using weightings derived from GWIS and GWEIS in UK Biobank (N = 23,092 and 99,057, respectively); and to assess whether the implementation of these alternative weightings to the *diathesis-stress* model would improve the evidence for GxE effects. First, we test the liability to a depression score explained by PRS weighted by the

effects estimated in UK Biobank GWIS and GWEIS in a sample of 4,919 unrelated white Caucasian participants from Generation Scotland. Then, we investigate the contribution of each *diathesis* in a full model accounting for the additive effects of SLE, PRS_{MDD}, PRS_{SS}, and PRS_{GxE} together. Finally, we compare the detected GxE effects with reported SLE in a *diathesis-stress* framework replacing the *diathesis-stress* interaction term by $SLExPRS_{MDD}$, $SLExPRS_{SS}$, or $SLExPRS_{GxE}$. In all three scenarios the main *diathesis* term in the *diathesis-stress* model is fixed to the best PRS_{MDD}[431] weighted by the Psychiatric Genomics Consortium (PGC) MDD GWAS[169]. Significant interactions are further investigated by level of exposure. By stratifying by sex we seek to explore differences between women and men in stress response underlying the aetiology of depression. We show an increased risk of depression not account for additive main contributions to MDD from variants displaying stress sensitivity and stress response effects. At current discovery sample sizes, the best estimates of a GxE effect come from testing the $SLExPRS_{MDD}$ interaction. However, it is possible that alternative implementations could achieve better prediction as sample sizes increase in line with the improvement of prediction seen using the main additive effects alone.

6.4 Materials and methods

6.4.1 Cohort description

This study is conducted in data from Generation Scotland (GS), a family-based population cohort representative of the population of Scotland[305]. GS is a Biobank of biological samples, health and lifestyle data established through a multi-institutional, cross-disciplinary collaboration. Blood and salivary DNA samples from donors across Scotland ware obtained at baseline, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. The Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry[413] was used to generate genome-wide genotype data. Details for DNA extraction and genotyping have been described elsewhere[307,414]. Invitation to take part in a follow-up mental health study (Stratifying Resilience and Depression Longitudinally, STRADL)[415] designed to investigate the aetiology of depression were sent to 21,525 eligible participants who gave consent for re-contact[415]. From those participants who agreed to provide new updated questionnaires to the follow-up, 8,541 respondents reported all the required measures for our study. These cover a wide range of psychiatric symptoms and stressful life events (SLE; see assessments below). We removed samples duplicated, population outliers not recorded as “white British” or outliers detected by principal component analysis (mainly non-Caucasians and Italian ancestry subgroup), participants who had a diagnosis of bipolar disorder, and samples with sex discrepancies or more than 2% missing genotype data. SNPs with a call rate < 98%, deviation from Hardy-Weinberg Equilibrium ($p < 1 \times 10^{-6}$), or a minor allele frequency < 0.01 were excluded from the analysis. Finally, participants were filtered by degree of relatedness (pi-hat < 0.05) maximising retention of those individuals reporting higher number of SLE. The final sample consisted of 4,919 unrelated individuals of white-European ancestry and 560,351 genotyped SNPs. The full cohort dataset (N = 4,919) was split by sex in two additional datasets: women (N = 2,990; mean age at 56.1) and men (N = 1,929; mean age at 58.7). Description of this study cohort has been previously published[169,431]. Further

details on the cohort profile of GS and STRADL are detailed elsewhere[124,304-306,415]. Ethical approvals were obtained from the Tayside Committee on Medical Research Ethics on behalf of the NHS (reference 05/s1401/89) and all data is available to researchers on application to the Generation Scotland Access Committee (access@generationscotland.org). All participants provided written consent.

6.4.2 Assessment of depressive symptoms

A self-administered 28-item scaled version of The General Health Questionnaire (GHQ)[416,417] was used to assess symptoms of depression. GHQ is a validated psychometric screening tool to detect current depressive symptoms[197], among other psychiatric conditions, and designed to identify whether the mental state of a participant at the time of reporting has changed over the preceding two weeks from their typical state. The questionnaire is composed of subscales of symptoms linked to depression (i.e. severe depression, anxiety, social dysfunction and somatic symptoms) highly correlated[418]. Thus, an overall general factor of depression is implied[419]. Participants used the Likert scoring to rate the severity of symptoms (e.g. *Have you recently felt that life is not worth living?* "Not at all", "No more than usual", "Rather more than usual", "Much more than usual"). The Likert scoring has the potential advantage to be more sensitive to psychopathology changes as response to stress in participants with chronic conditions (i.e. whose current symptoms do not differ from their typical state)[420], and to provide a wider and smoother distribution than the summation scoring of number of symptoms[197]. To approximate to a normal distribution and reduce the effect of positive skew, the score was log-transformed to generate the final depression score. To validate the usefulness of our score as a measure of depression, we used self-reported questionnaire data on the Composite International Diagnostic Interview–Short Form, which diagnoses lifetime history of MDD according to DSM-IV criteria[42]. Depression score was scaled to a mean of 0 for a better interpretation of illustrations in **Figure 6.3b**.

6.4.3 Stressful life events (SLE)

A measure of recent SLE was constructed from a self-reported brief life events questionnaire based on the 12-item List of Threatening Experiences[99]. Participants reported SLE experienced over the preceding 6 months. This is a reliable psychometric screening tool to measure psychological stress with considerable long-term contextual effects (e.g. *Over last 6 months, did a parent, spouse/partner, child, brother or sister of yours die?*)[422,423]. A measure reflecting the total number of SLE reported was constructed by combining the number of "yes" responses.

6.4.4 Polygenic profiling

Polygenic risk scores (PRS) aggregate the number of risk alleles carried by an individual weighted by their allelic effects estimated in a genome-wide study from an independent discovery sample. We created four sets of PRS using summary statistics from four different studies: two Psychiatric Genetic Consortium (PGC) GWAS for MDD (stage 1: 9,240 cases and 9,519 controls; and updated version: 50,455 cases and 105,411 controls)[147,169]; a Genome-Wide Interaction Study (GWIS) conducted on 23,092 unrelated white British participants from UK Biobank[378]; and a Genome-Wide by Environment Interaction Study (GWEIS) of depressive symptoms and psychological stress involving 99,057 unrelated white British individuals from UK Biobank[476]. The last three studies had Generation Scotland participants excluded. PRS were generated by PRSice-2[313] and calculated from individual-level genotype data on each participant (N = 4,919). PRSice-2 removed strand-ambiguous SNPs and obtained the most significant independent SNPs in approximate linkage equilibrium by accounting for linkage disequilibrium (LD) among SNPs (LD-clump-based pruning: r^2 = 0.1, within a 10Mb window). First, two sets of PRS_{MDD} were generated using summary statistics from the PGC-MDD GWAS and weighted by main additive effects for MDD[147,169]. A third set of PRS_{SS} was generated using summary statistics from the UK Biobank GWIS and weighted by stress-sensitivity effects[378]. Finally, a fourth set of PRS_{GxE} was generated using summary statistics from the UK Biobank GWEIS and weighted by stress-response (GxE) effects[427]. PRS were generated for seven

p thresholds (*p* threshold < 1 x 10^{-5}, < 0.001, < 0.01, < 0.05, < 0.1, < 0.5, <=1) determined by each discovery sample. PRS_{MDD} , PRS_{SS} and PRS_{GxE} represent three distinct genetic components or metrics of *diathesis.*

PGC GWAS results are available at https://www.med.unc.edu/pgc/results-and-downloads. Statistics excluding GS participants may be obtained under request. GWIS results are fully available[378]. Requests for the GWEIS results should be made at the Institute of Genetics and Molecular Medicine, Edinburgh, by contacting Dr. Pippa Thomson (pippa.thomson@ed.ac.uk).

6.4.5 PRS models

For each dataset (full sample, women and men), we constructed a genetic relatedness matrix (GRM) to fit into mixed linear models implemented in GCTA 1.26.0[176]. First, we tested the effect of both PRS_{SS} and PRS_{GxE} predicting both the depression score and reported SLE using mixed linear models as follow:

$$Depression = \beta_0 + \beta_1 PRS_{SS} + GRM + Covariates$$

$$Depression = \beta_0 + \beta_1 PRS_{GxE} + GRM + Covariates$$

$$Reported\ SLE = \beta_0 + \beta_1 PRS_{SS} + GRM + Covariates$$

$$Reported\ SLE = \beta_0 + \beta_1 PRS_{GxE} + GRM + Covariates$$

Covariates include: age, age^2, sex, age-by-sex and age^2-by-sex interactions and 20 principal components. The significance of β_1 (i.e. the effect of PRS) was assessed using a one-sided Wald-test. To account for multiple testing correction, we established a threshold for significance at *p* = 3.57 x 10^{-3} after correcting for seven thresholds and two different *diathesis* effects tested (14 tests).

We assessed the effect of PRS_{MDD} to predict depression score weighting by the past stage 1 PGC-MDD GWAS (referred as $PRS_{MDD-stage1}$)[147] to compare its performance against the effects of PRS_{MDD} weighting by the

largest sample size previously reported[169,431]. The effect of PRS_{MDD} predicting the depression score and reported SLE using the largest PGC-MDD GWAS, as well as the effect of the total number of reported SLE predicting the depression score, are published elsewhere[431].

6.4.6 A multi-PRS model

To investigate the contribution of each *diathesis* we combined both SLE, PRS_{MDD}, PRS_{SS} and PRS_{MDD} into a model in the full sample and stratifying by sex as follows:

$$Depression\ score = \beta_0 + \beta_1 SLE + \beta_2 PRS_{MDD} + \beta_3 PRS_{SS} + \beta_4 PRS_{GxE} + GRM + Covariates$$

Covariates adjusted were the same detailed above for the *diathesis* models. The variance in depression score explained by all parameters and the significance of each effect coefficient (β_1, β_2, β_3 and β_4) was estimated using a one-sided Wald test.

6.4.7 Diathesis-stress models

In our set of GS individuals, PRS_{MDD} generated at the PGC-MDD GWAS *p* threshold = 0.1 ($PRS_{MDD\_0.1}$) was previously reported as the best predictor of the depression score[169,431]. To test the different *diathesis-stress* models for depression, we used $PRS_{MDD\_0.1}$ to keep the main *diathesis* effect unchangeable across studies. Following previous implementations of the *diathesis-stress* model[169,431], we fitted $PRS_{MDD\_0.1}$ into the main term of the model (i.e. the main *diathesis* effect) and replaced the interactive product term (i.e. the *diathesis-stress* interaction effect) with either $SLExPRS_{MDD}$, $SLExPRS_{SS}$ or PRS_{GxE}. $SLExPRS_{MDD}$ were tested using both PRS_{MDD} and PRS_{MDD_stage1} to assess how increasing the discovery sample size influence the final output. The tested models were as follows:

$$Model\ 1: Depression = \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{MDD} + GRM + Cov.$$

$$Model\ 2: Depression = \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{MDD\_stage1} + GRM + Cov.$$

$$Model\ 3: Depression = \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{SS} + GRM + Cova.$$

$$Model\ 4: Depression = \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{GxE} + GRM + Cov.$$

Covariates (*Cov.*) include: age, age^2, sex, age-by-sex and age^2-by-sex interactions and 20 principal components. In order to adjust for covariate interactions[424], the covariates were previously regressed out from all PRS and SLE and the residuals standardized to a mean of 0 and standard deviation of 1. This allows to account for potential confounding influences on $SLExPRS$ interactions[424]. Therefore, *PRS* and *SLE* represent the residuals of these metrics. Sex, age-by-sex and age^2-by-sex were not fit when stratifying by sex. All models were stratified by sex. All parameters from the models and the variance in depression score explained by the interactive component were estimated. The significance of β_3 (i.e. the effect of the *diathesis-stress* interaction term) was nominally assessed using a two-sided Wald-test. To account for multiple testing correction, a Bonferroni's adjustment correcting for seven thresholds and three different *diathesis-stress* effects tested (21 tests) was used to establish a threshold for significance at $p = 2.38 \times 10^{-3}$.

6.4.8 Differences across estimates

Z-scores derived from the effect estimates (β) and standard errors (SE) were calculated and compared to the standard normal distribution (α = 0.05, one-tailed) to assess significant differences between pairs of estimated effects (either within a model or across models). For example, z-score is calculated to investigate differences between a *diathesis* effect in women and men as follows:

$$\mathrm{Z-score} = \frac{\beta_{women} - \beta_{men}}{\sqrt{SE(\beta_{women})^2 + SE(\beta_{men})^2}}$$

6.4.9 Examination of PRS_{SS} effects by levels of exposure

For the significant $SLExPRS$ detected (i.e. $SLExPRS_{SS}$), the number of SLE reported were categorized to examine the PRS effect at different levels of SLE. To retain a large enough sample size for each category and thus allow

meaningful statistical comparison, three categories were defined: “none” for 0 SLE reported, “low” for 1 or 2 SLE reported, and “high” for 3 or more SLE reported. We explored the *diathesis* effect of PRS_{SS} on the depression score at each category of SLE reported using the threshold where we detected a significant interaction in men surviving multiple testing (PRS_{SS} *p*-threshold = 0.01) and using the best interactive threshold at full sample (PRS_{SS} *p*-threshold = 1 x 10^{-5}). For comparison, we examine the effect in the full sample, men and women. The method of least squares to test the PRS_{SS} effect on the depression score was applied in each category of SLE reported. The prior significance threshold $p = 2.38 \times 10^{-3}$ to assess the *diathesis-stress* interaction was further adjusted for the 3 SLE categories tested resulting in a threshold for robust significance set at $p = 7.94 \times 10^{-4}$ in this analysis.

6.5 Results

6.5.1 Do the PRS_{SS} and PRS_{GxE} predict depression score or number of SLE?

The depression score significantly predicted lifetime history of MDD screened with the Composite International Diagnostic Interview–Short Form in GS respondents (odd ratio = 1.91, 95% confidence intervals 1.80-2.02, *p*-value = 1.55 x 10^{-102}, N = 8,994), with those respondents in the top versus bottom depression score decile with odds ratio of 3.8 of having a lifetime history of MDD.

Previously, PRS_{MDD} was reported to explain 0.64% of the depression score variation in our sample (β = 0.080, s.e. = 0.014, *p* = 7.53 x 10^{-9})[431], whereas PRS_{MDD} weighted by the effects derived at a lower sample size using past PGC-MDD1 GWAS (9,240 cases and 9,519 controls; referred as $PRS_{MDD-stage1}$) explained 0.21% of variation (β = 0.046, s.e. = 0.014, *p* = 5.99 x 10^{-4})[147]. Similarly we tested both PRS_{SS} and PRS_{GxE} ; both PRS significantly predicted the depression score in the full target sample (N = 4,919) explaining a maximum variance of 0.24% and 0.15%, respectively (PRS_{SS}: β = 0.049, s.e. = 0.014, *p* = 2.63 x 10^{-4}; PRS_{GxE}: β = 0.038, s.e. = 0.014, *p* = 3.28 x 10^{-3}; **Figure 6.1a**). After stratifying the target cohort by sex, predictions were nominally significant in both women and men for both PRS_{SS} and PRS_{GxE} (**Figure 6.1a**). However, only prediction of PRS_{SS} in men remained significant after multiple testing corrections (β = 0.076, s.e. = 0.022, *p* = 2.09 x 10^{-4}; see **Figure 6.1a**). Similar pattern was seen with $PRS_{MDD-stage1}$ (men: β = 0.064, s.e. = 0.022, R^2 = 0.42%, *p* = 1.52 x 10^{-3}; women: β = 0.037, s.e. = 0.019, R^2 = 0.13% s, *p* = 0.025). The variance explained by PRS_{SS} in men (maximum R^2 = 0.59%) was nominally greater than in women (maximum R^2 = 0.13%), but not significantly (*p* = 0.068 at the best *p*-threshold 0.1).

The effects and variance in SLE explained by both PRS_{MDD} has been previously published elsewhere[431]. PRS_{MDD} significantly explained 0.23% of variation in SLE reported (p = 3.64 x 10^{-4}) in the full sample, and 0.41% (p = 2.22 x 10^{-3}) and 0.22% (p = 4.86 x 10^{-3}) in men and women, respectively[431]. Unlike the PRS_{MDD}, neither PRS_{SS} or PRS_{GxE} predicted the variation in the number of SLE reported (PRS_{SS}: full sample: R^2 = 0.02%, p = 0.079; men: R^2 = 0.05%, p = 0.060; women: R^2 = 0.01%, p = 0.204; and PRS_{GxE}: full sample: R^2 = 0.01%, p = 0.879; men: R^2 = 0.08%, p = 0.973; women: R^2 = 0.01%, p = 0.243; **Figure 6.1b**).

6.5.2 A multi-PRS model for depression score

In each sample, we combined the measure of self-reported SLE with the best predictors from all three indicators of *diathesis* (i.e. PRS_{MDD}, PRS_{SS} and PRS_{GxE}) in a multi-PRS model of depression score (see **Table 6.1** and **Figure 6.2**). All three PRS had a significant contribution to depression score in the full sample (PRS_{MDD}: R^2 = 0.58%, p = 7.38 x 10^{-8}; PRS_{SS}: R^2 = 0.19%, p = 7.38 x 10^{-4}; PRS_{GxE}: R^2 = 0.08%, p = 0.019), accounting for a total amount of variance in depression score of 0.86%; and in men (PRS_{MDD}: R^2 = 0.47%, p = 7.46 x 10^{-4}; PRS_{SS}: R^2 = 0.49%, p = 5.36 x 10^{-4}; PRS_{GxE}: R^2 = 0.13%, p = 0.049), accounting for 1.08% of depression score; but not in women, as the contribution of PRS_{GxE} effect was not significant (PRS_{MDD}: R^2 = 0.68%, p = 2.91 x 10^{-6}; PRS_{SS}: R^2 = 0.11%, p = 0.032; PRS_{GxE}: R^2 = 0.06%, p = 0.081), with the PRS effects accounting for 0.86% of depression score (**Table 6.1** and **Figure 6.2**). In the full sample, the genetic contributions to depression score attributable to additive main effects of MDD were higher than those attributable to effects of a proxy for stress-sensitivity (p = 0.05) or attributable to GxE effects (p = 7.59 x 10^{-3}). Similar pattern was seen in women, with the contribution of PRS_{MDD} higher than the contribution of PRS_{SS} (p = 0.03) or PRS_{GxE} (p = 0.01). However, in men, the estimated contribution of PRS_{MDD} to depression score was not significantly greater than the contribution of PRS_{SS} (p = 0.46) or PRS_{GxE} (p = 0.13; **Table 6.1** and **Figure 6.2**). Findings suggest that in men, weightings derived from the GWIS and GWEIS non-attributable

to additive main effects of MDD at the current discovery sample sizes may be as relevant as the effects detected in standard GWAS of MDD to explain variation in risk of depression. Women have a higher prevalence of MDD than men; therefore, standard GWAS may be bias towards female-specific additive effects. However, the estimated PRS effects were not different between women and men (PRS_{MDD}: *p* = 0.329; PRS_{SS}: *p* = 0.087; PRS_{GxE}: *p* = 0.476). There was no difference in the effects of reported SLE across sexes (*p* = 0.186). These results require of further replication in independent studies and larger sample sizes to draw robust conclusions. However, they suggest a significant genetic contribution from genetic stress response mechanisms not detected as additive main effects by GWAS and emphasize the potential of alternative genome-wide approaches.

6.5.3 GxE effects, PRS_{SS} and PRS_{GxE}, to test the diathesis-stress model

We investigated the presence of a GxE effect with reported SLE in a diathesis-stress framework replacing the *diathesis-stress* interaction term by $SLExPRS_{MDD}$, $SLExPRS_{SS}$, or $SLExPRS_{GxE}$. In all *diathesis-stress* models, the GxE effect was considerably smaller than the main additive effect (operationalized as $PRS_{MDD\_0.1}$ at the best predictive p-threshold 0.1 across all models; see **Table 6.2**). We did not detect significant GxE effects of either $SLExPRS_{SS}$ (R^2 =0.04%, p = 0.132) nor $SLExPRS_{GxE}$ (R^2 =0.02%, *p* = 0.330) in the full sample (**Table 6.2** and **Figure 6.3a**). The only prediction surviving Bonferroni correction was for a $SLExPRS_{SS}$ interaction in men (**Figure 6.3**). This significantly contributed to the prediction of the depression score at GWIS *p*-threshold = 0.01 in men (R^2 = 0.42%, β = 0.065, s.e. = 0.021, *p* = 2.1 x 10^{-3}; see **Table 6.2**) and had a significantly greater effect in men than in women (*p* = 1.08 x 10^{-3}). Estimation of a $SLExPRS_{SS}xsex$ in the full sample supported this sex-specific GxE effect (*p* = 1.00 x 10^{-3}; PRS_{SS} *p*-threshold = 0.01).

Overall, the full *diathesis-stress* models predicted depression score better in women than in men, due to an increased ability of the number of SLE reported to predict liability, regardless of the GxE weightings (**Table 6.2**).

To assess whether the evidence of GxE effects would likely increase with larger discovery sample sizes we tested the $SLExPRS_{MDD}$ interaction effect using PRS_{MDD} weighted by either PGC MDD2 GWAS (50,455 cases and 105,411 controls)[169] or PGC MDD1 GWAS (9,240 cases and 9,519 controls; referred as $SLExPRS_{MDD-stage1}$)[147]. In the full sample, we detected nominally significant effects of $SLExPRS_{MDD}$ (R^2 = 0.08%, p = 0.044), as reported previously[431], but not of $SLExPRS_{MDD-stage1}$ (R^2 = 0.03%, p = 0.201; **Table 6.2** and **Figure 6.3a**), showing the relevance of large enough discovery sample size to detect GxE effects. Nominally significant effects were also detected for $SLExPRS_{MDD}$ in women (p = 0.016). Our results suggest that the best *diathesis-stress* indicator to test for the presence of GxE effects across the entire population with the current discovery sample sizes available is the $SLExPRS_{MDD}$. However, as seen in comparison to results using $SLExPRS_{MDD\_stage1}$, using alternative weightings could provide better results with large enough GWIS and discovery samples.

6.5.4 Examination of the $SLExPRS_{SS}$ effect by SLE categories

Given the detection of a significant $SLExPRS_{SS}$ interaction in men (p-threshold = 0.01), we further examined the effect of PRS_{SS} on depression score at different levels of stress using three categories of SLE reported (i.e., “none”, “low” or “high” number of SLE reported; see **Table 6.3** and **Figure 6.3b**). For comparison, we also assessed such PRS_{SS} effect (PRS_{SS} p-threshold = 0.01) in all three subsamples (**Table 6.3**). The results suggest that the depression score is higher the larger the number of SLE reported. In men, the effect of SLE was modulated by PRS_{SS}. We found that high PRS_{SS} (i.e. a high genetic load for a stress-sensitivity proxy) significantly increased depression score in men who reported a high number of SLE (N = 285, β = 0.246, s.e. = 0.065, p = 1.84 x 10^{-4}; **Table 6.3**). This effect was also nominally significant in the full

sample (N = 1833, β = 0.079, s.e. = 0.04, *p* = 0.047), but not in women (N = 490, β = -0.019, s.e. = 0.05, *p* = 0.71).

In addition, we investigate the *diathesis* effect of PRS_{SS} by levels of SLE using the threshold displaying the strongest $SLExPRS_{SS}$ interaction in the full sample (*p* = 0.13, PRS_{SS} *p* threshold = 1 x 10^{-5}). This includes only 5 top independent SNPs from the GWIS of a proxy for stress-sensitivity. The *diathesis* effect conferred by these 5 SNPs increased the risk of depression in participants who reported high number of SLE (N = 775, β = 0.108, s.e. = 0.04, *p* = 7.37 x 10^{-3}; see **Table 6.3** and **Figure 6.3b**). However, this finding did not survive multiple testing correction. This effect was also nominally detected in women (N = 490, β = 0.111, s.e. = 0.051, *p* = 0.03), but not in men (N = 285, β = 0.105, s.e. = 0.066, *p* = 0.114) maybe due to their lower sample size and the corresponding reduced power of considering the effect from just 5 SNPs.

Table 6.1 A full *diathesis* single-model for depression score

	FULL COHORT				MEN				WOMEN			
Parameter	**SLE**	**PRS_{MDD}**	**PRS_{SS}**	**PRS_{GxE}**	**SLE**	**PRS_{MDD}**	**PRS_{SS}**	**PRS_{GxE}**	**SLE**	**PRS_{MDD}**	**PRS_{SS}**	**PRS_{GxE}**
***best p*-threshold**	-	0.1	0.1	0.1	-	0.1	0.1	0.1	-	0.1	0.05	0.1
beta	0.210	0.076	0.044	0.029	0.195	0.069	0.072	0.035	0.220	0.082	0.033	0.026
s.e.	0.014	0.014	0.014	0.014	0.021	0.022	0.022	0.021	0.018	0.018	0.018	0.018
z-score	15.12	5.50	3.18	2.07	9.13	3.18	3.27	1.66	12.07	4.53	1.84	1.40
***p* value**	5.96×10^{-52}	1.83×10^{-8}	7.38×10^{-4}	0.019	3.43×10^{-20}	7.46×10^{-4}	5.36×10^{-4}	0.049	7.15×10^{-34}	2.91×10^{-6}	0.032	0.081
r^2 (%)	4.409	0.585	0.195	0.083	3.802	0.466	0.492	0.126	4.836	0.681	0.113	0.065
Genetics r^2 (%)		0.862				1.085				0.859		
Total r^2 (%)	5.271				4.887				5.695			

Model: $Depression\ score \sim \beta_0 + \beta_1 SLE + \beta_2 PRS_{MDD} + \beta_3 PRS_{SS} + \beta_4 PRS_{GxE} + GRM + Covariates$.

Each parameter selected using the best *p*-threshold predicting depression score in a model without the other parameters (see **Figure 6.1**). Significance of each 'Parameter' is assessed using a one-sided Wald-test. r^2 shows the percentage of depression score explained by each parameter; Genetics: combined *diathesis* scores; Total: including the effect of SLE (reported stressful life events).

Table 6.2 Comparison between *diathesis-stress* models across different weightings: main additive, stress-sensitivity or GxE effects

	MODEL 1: PRS_{MDD}			MODEL 2: PRS_{MDD} (PGC MDD1)			MODEL 3: PRS_{SS}			MODEL 4: PRS_{GxE}		
FULL COHORT												
Parameter	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD_stage1}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{SS}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**PRS_{GxE}**
***p*-threshold**	0.1	-	0.01	0.1	-	$1x10^{-3}$	0.1	-	$1x10^{-5}$	0.1	-	0.5
beta	0.075	0.218	0.028	0.071	0.219	-0.018	0.078	0.217	0.021	0.078	0.218	-0.013
s.e.	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
z-score	5.38	15.90	2.01	5.15	15.95	-1.28	5.70	15.79	1.50	5.68	15.90	-0.97
***p* value**	$3.76x10^{-8}$	$3.41x10^{-8}$	0.044	$1.29x10^{-7}$	$1.44x10^{-57}$	0.201	$6.09x10^{-9}$	$1.71x10^{-56}$	0.132	$6.76x10^{-9}$	$3.04x10^{-57}$	0.330
r^2 (%)	0.557	4.763	0.078	0.502	4.796	0.031	0.613	4.715	0.043	0.609	4.764	0.018
Total r^2 (%)		5.398			5.329			5.37			5.391	
MEN												
Parameter	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD_stage1}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{SS}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{GxE}$**
***p*-threshold**	0.1	-	0.1	0.1	-	0.1	0.1	-	0.01	0.1	-	0.05
beta	0.081	0.202	0.039	0.072	0.203	-0.018	0.071	0.205	0.065	0.072	0.201	-0.018
s.e.	0.022	0.021	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021
z-score	3.70	9.55	1.80	3.30	9.59	-0.86	3.34	9.71	3.08	3.41	9.52	-0.84
***p* value**	$1.07x10^{-4}$	$6.57x10^{-22}$	0.072	$4.75x10^{-4}$	$4.25x10^{-22}$	0.388	$4.19x10^{-4}$	$1.43x10^{-22}$	$2.10x10^{-3}$	$3.31x10^{-4}$	$9.01x10^{-22}$	0.401
r^2 (%)	0.65	4.07	0.153	0.512	4.11	0.034	0.486	4.195	0.424	0.523	4.054	0.032
Total r^2 (%)		4.873			4.656			5.106			4.609	
WOMEN												
Parameter	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{MDD_stage1}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{SS}$**	**$PRS_{MDD\_0.1}$**	**SLE**	**SLExPRS$_{GxE}$**
***p*-threshold**	0.1	-	$1x10^{-5}$	0.1	-	$1x10^{-3}$	0.1	-	0.05	0.1	-	0.5
beta	0.071	0.227	0.044	0.069	0.228	-0.033	0.082	0.227	-0.028	0.081	0.227	-0.025
s.e.	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
z-score	3.96	12.63	2.41	3.87	12.65	-1.82	4.55	12.60	-1.56	4.53	12.64	-1.39
***p* value**	$3.79x10^{-5}$	$7.37x10^{-37}$	0.016	$5.40x10^{-5}$	$5.37x10^{-37}$	0.069	$2.66x10^{-6}$	$1.05x10^{-36}$	0.118	$3.01x10^{-6}$	$6.19x10^{-37}$	0.165
r^2 (%)	0.508	5.163	0.189	0.479	5.189	0.108	0.683	5.137	0.08	0.664	5.171	0.063
Total r^2 (%)		5.861			5.776			5.899			5.898	

Model 1: $Depression\ score \sim \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{MDD} + GRM + Covariates.$

Model 2: $Depression\ score \sim \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{MDD\_stage1} + GRM + Covariates.$

Model 3: $Depression\ score \sim \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{SS} + GRM + Covariates.$

Model 4: $Depression\ score \sim \beta_0 + \beta_1 PRS_{MDD\_0.1} + \beta_2 SLE + \beta_3 SLExPRS_{GxE} + GRM + Covariates.$

PRSMDD_0.1, selected at the best p-threshold predicting depression score, was unchangeable across models. SLE: reported stressful life events. *p*-value reflects the significance of each 'Parameter' in the model assessed using a one-sided, for the effect of $PRS_{MDD\_0.1}$ and SLE, and a two-sided Wald-test for the effect of interactions. r^2 reflects the percentage of depression score explained by each parameter.

Table 6.3 PRS_{SS} effect under 3 levels of SLE reported

diathesis	Stress-sensitivity effect derived from GWIS at p value threshold = 0.01								
Sample	**FULL COHORT**			**MEN**			**WOMEN**		
SLE category	**none**	**low**	**high**	**none**	**low**	**high**	**none**	**low**	**high**
N	1833	2311	775	792	852	**285**	1041	1459	490
Effect	0.0122	0.0218	0.0791	-0.0018	0.045	**0.2465**	0.0219	0.0086	-0.0186
CI (95%)	-0.030, 0.055	-0.019, 0.063	0.001, 0.157	-0.068, 0.064	-0.019, 0.109	**0.119, 0.375**	-0.034, 0.078	-0.045, 0.062	-0.117, 0.079
s.e.	0.022	0.021	0.04	0.034	0.033	**0.065**	0.029	0.027	0.05
t	0.562	1.038	1.986	-0.055	1.378	**3.79**	0.764	0.313	-0.372
***p* value**	0.574	0.299	0.047	0.956	0.169	**1.84×10^{-4}**	0.445	0.754	0.71
diathesis	**Stress-sensitivity effect derived from GWIS at p value threshold = 1×10^{-5}**								
Sample	**FULL COHORT**			**MEN**			**WOMEN**		
SLE category	**none**	**low**	**high**	**none**	**low**	**high**	**none**	**low**	**high**
N	1833	2311	775	792	852	285	1041	1459	490
Effect	0.0036	0.0152	0.1083	-0.0227	0.0057	0.1044	0.0242	0.021	0.1108
CI (95%)	-0.039, 0.046	-0.026, 0.057	0.029, 0.187	-0.086, 0.040	-0.057, 0.069	-0.025, 0.234	-0.033, 0.082	-0.033, 0.075	0.010, 0.211
s.e.	0.022	0.021	0.04	0.032	0.032	0.066	0.029	0.028	0.051
t	0.168	0.723	2.687	-0.707	0.176	1.585	0.827	0.761	2.17
***p* value**	0.867	0.47	7.37×10^{-3}	0.48	0.86	0.114	0.408	0.447	0.03

SLE categories (amount of SLE reported): "none" = 0 SLE, "low" = 1 or 2 SLE, and "high" = 3 or more SLE reported. N shows the sample size of each SLE category. *p* value reflects significance of the diathesis effect on depression score in individuals within each SLE category.

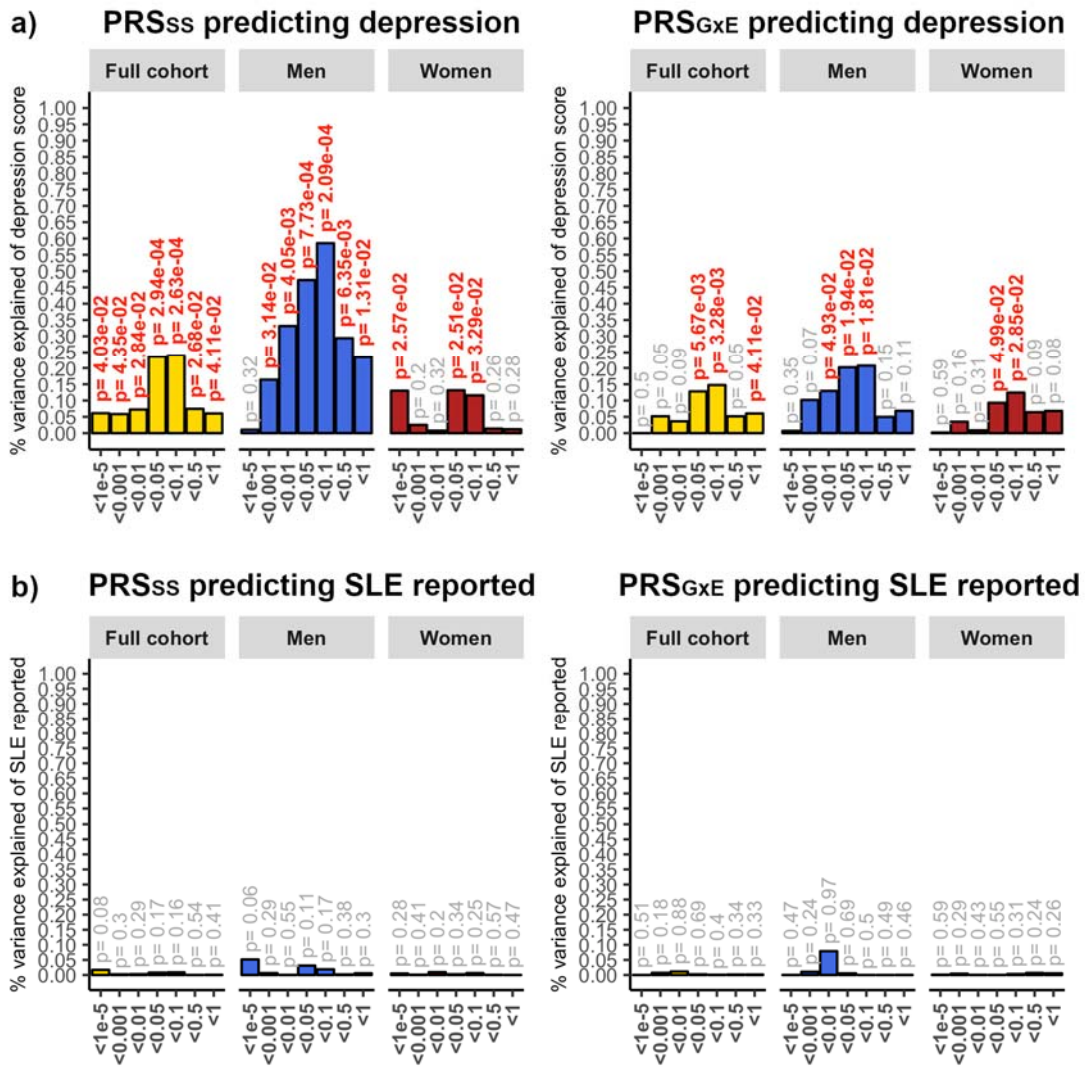


Figure 6.1 Association between PRS_{SS} and PRS_{GxE} with depression score. *PRS were generated using external weights from analyses using UK Biobank participants with the exclusion of Generation Scotland participants and their relatives* ***a)*** PRS_{SS} *were generated using weights from a Genome-Wide Interaction Study (GWIS) of a proxy for stress-sensitivity* ***b)*** PRS_{GxE} *were generated using GxE effects from a Genome-Wide by Environment Interaction Study (GWEIS) of depressive symptoms and common stressful life events. Y-axis represents the % of variance of depression score explained by each PRS (main effects, one-sided Wald-tests). Red numbers show significant p-values. PRS were constructed at 7 GWIS/GWEIS p-threshold levels. Full cohort (yellow) was split into men (blue) and women (red).*

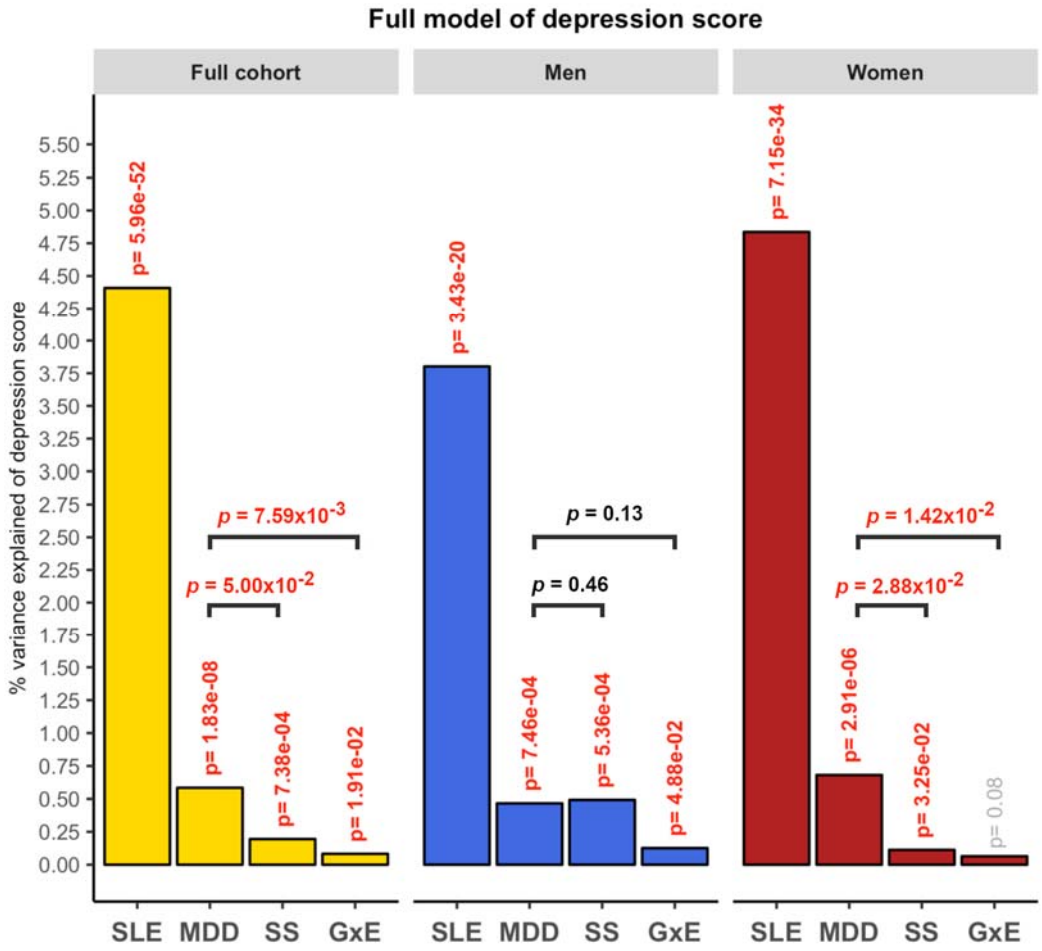


Figure 6.2 The multi-PRS model for depression score across samples. *SLE: the total number of SLE reported.* PRS_{MDD} *weighted by the effects from the Psychiatric Genetic Consortium MDD GWAS.* PRS_{SS} *weighted by the effects from a Genome-Wide Interaction Study (GWIS) of a proxy for stress-sensitivity.* PRS_{GxE} *weighted by the GxE effects from Genome-Wide by Environment Interaction Study (GWEIS) of depressive symptoms and common stressful life events. Y-axis represents the % of variance of depression score explained by each parameter from the multi-PRS model. Red numbers show significant p-values (one-sided Wald-tests). Differences between the estimated effect of* PRS_{MDD} *and the other genetic response diathesis are shown. Full cohort (yellow) was split into men (blue) and women (red).*

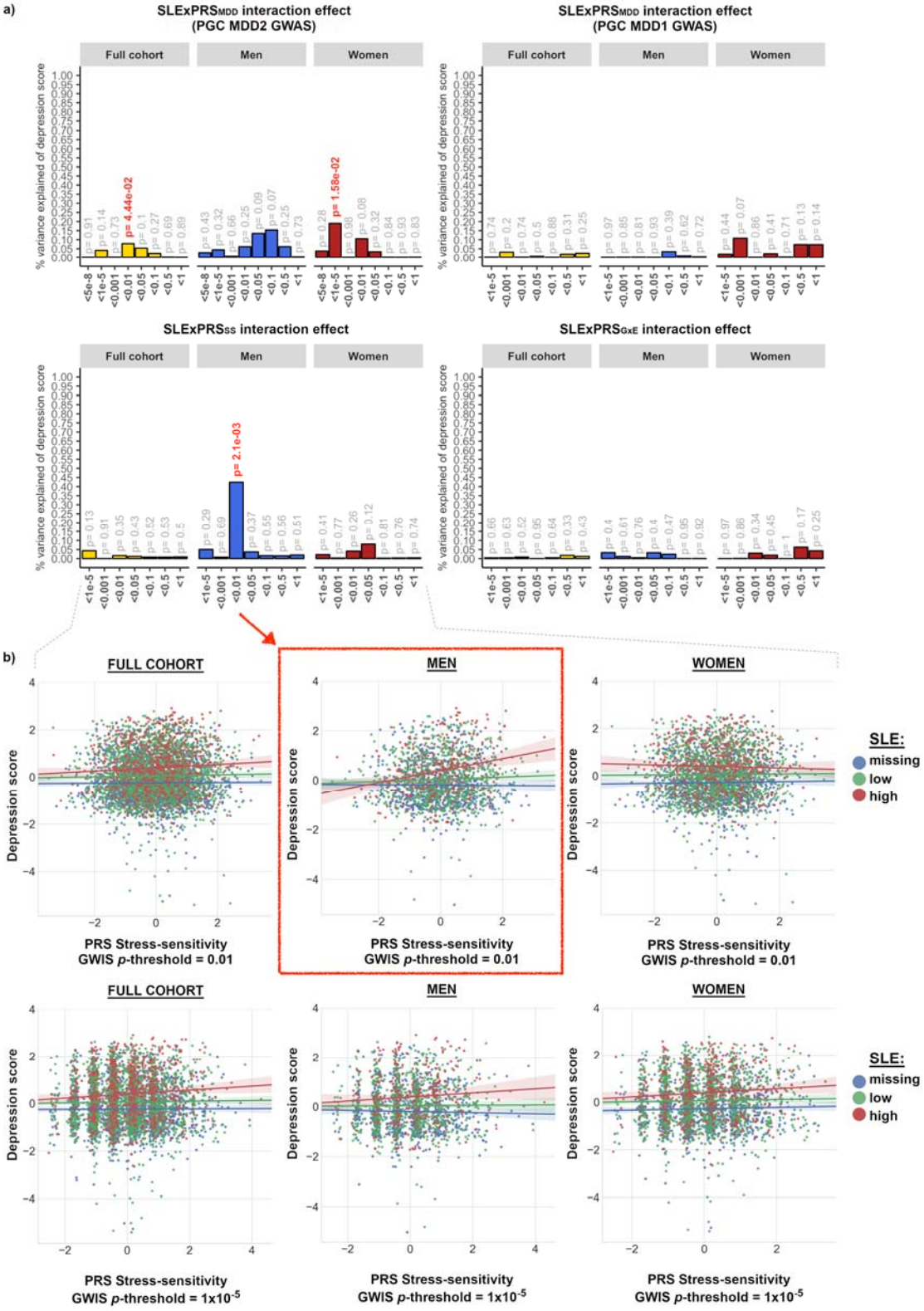

a)
SLExPRSMDD interaction effect (PGC MDD2 GWAS)
SLExPRSMDD interaction effect (PGC MDD1 GWAS)
SLExPRSSS interaction effect
SLExPRSGxE interaction effect
Full cohort
Men
Women
% variance explained of depression score
p= 4.44e-02
p= 1.58e-02
p= 2.1e-03
b)
FULL COHORT
MEN
WOMEN
Depression score
PRS Stress-sensitivity
GWIS p-threshold = 0.01
GWIS p-threshold = 1x10-5
SLE:
missing
low
high

Figure 6.3 a) Association between *diathesis-stress* interaction (GxE) effects and depression score. *The interaction effect was tested as:* $SLExPRS_{MDD}$ *(top row) using* PRS_{MDD} *weighted by either the PGC GWAS for MDD (50,455 cases and 105,411 controls; top-left) or a past stage 1 PGC GWAS for MDD (9,240 cases and 9,519 controls; top-right);* $SLExPRS_{SS}$ *(bottom-left);* $SLExPRS_{GxE}$ *(bottom-right).* PRS_{SS} *weighted by the effects from a Genome-Wide Interaction Study (GWIS) of a proxy for stress-sensitivity.* PRS_{GxE} *weighted by the GxE effects from Genome-Wide by Environment Interaction Study (GWEIS) of depressive symptoms and common stressful life events. Discovery samples had Generation Scotland participants excluded. SLE: the total number of SLE reported. Results represent variation of depression score explained by the diathesis-stress interaction term fitted in linear mixed models to empirically test the diathesis-stress model for depression. The same main effect of* PRS_{MDD} *as additive main diathesis term (i.e. genetic additive main contributions to major depressive disorder) was taken into account across all models. Red numbers show nominally significant p-values. Full cohort (yellow) was split into men (blue) and women (red). PRS were generated at 7 GWIS/GWEIS p-threshold levels*. **b) Scatterplots of *diathesis-stress* interactions between PRS_{SS} and levels of SLE reported on the risk of depression score**. *Top row: GWIS p-threshold with which a significant interaction was detected in men (centre, red square). For comparison, we show results at the same p-threshold from the full sample (left) and women (right). Bottom row: GWIS p-threshold displaying the strongest interaction in the full sample (p = 1 x* 10^{-5}*). The number of SLE reported by each participant (dot) were categorized at 3 levels represented by colours. Blue: 0 SLE, "no stress", n = 1,833/1,041/792; green: 1 or 2 SLE, "low stress", n = 2,311/1,459/852; red: 3 or more SLE, "high stress", n = 775/490/285; in the full cohort/women/men, respectively. Y-axis reflects depression score standardized to mean of 0 and standard deviation of 1. Lines represent the increment of depression score under a certain degree of reported SLE dependent on the genetic predisposition (= diathesis) conferred by* PRS_{SS}*. See **Table 6.3** for further details.*

6.6 Discussion

The depressogenic effects of SLE may not act exclusively through interaction with known genetic risk factors of MDD, but also through the interplay with alternative genetic factors specific to sensitivity and response to stress[469-471]. The rational of previous studies assessing $SLExPRS$ interactions (i.e. conceptualized as the interaction term on a *diathesis-stress* model) relied on the untested assumption that those genetic variants that display an additive main effect for MDD in case/control GWAS, and that are used to construct the PRS as indicators of *diathesis,* are the same common genetic variants implicated in GxE. In contrast, this study combined the main additive contributions for MDD (PRS_{MDD}) with alternative interaction terms weighted by either genetic stress-sensitivity (PRS_{SS}) or stress-response (PRS_{GxE}) effects, while controlling for the direct effect of reported SLE. While using these weightings did not improve the results using the main additive weightings, it is possible that such alternative implementations could improve the testing of GxE effects with big enough discovery sample sizes. Our results support a genetic contribution to risk of depression from genetic responses to stress non-attributable to additive main contributions of MDD detected in current GWAS.

Both $PRS_{SS,}$ weighted by the effects estimated in a Genome-Wide Interaction Study (GWIS) of a proxy for stress-sensitivity[378], and PRS_{GxE}, weighted by GxE effects with a measure of common SLE derived from a Genome-Wide by Environment Interaction Study (GWEIS)[427], significantly predicted variation in the depression score (0.24% and 0.15%, respectively) in the full sample. This is in contrast to the 0.64% of the variance explained by PRS_{MDD} weighted by common additive main effects of MDD in this same sample, and reported elsewhere[431], but aligns with the variance explained by PRS_{MDD_stage1} (0.21%) weighted by past PGC-MDD1 GWAS at lower discovery sample size (N = 9,240 cases and 9,519 controls)[147]. Stratifying by sex, the effect of PRS_{SS} in men remained significant after multiple testing

correction (R^2 = 0.59%, p = 2.09 x 10^{-4}), with an overall explanatory power equivalent to the explanatory power of PRS_{MDD}[431]. Unlike PRS_{MDD}[431], neither PRS_{SS} nor PRS_{GxE} predicted variation in the number of SLE reported, suggesting that the implementation of such alternative weightings to test for GxE effects could reduce a potential confounding bias due to a gene-environment correlation effect[426,478]. Individuals vulnerable to MDD may have behavioural traits that increase their likelihood of experiencing a SLE[103,107] and those with the disease have a propensity to report more SLE than healthy controls[98]. Genetic sensitivity to stress and stress response may moderate the causal relationship between stress and depression. Evidence from life-course GxE approaches suggests that such mediation is dependent and modulated by previous cumulative interactions experienced over lifespan[479-481].

Jointly fitting the three PRS and SLE in a single model of depression score suggests that the PRS effects tested are relevant to depression (**Table 6.2** and **Figure 6.2**). However, in this multi-PRS model we found that the contribution of PRS_{MDD} to depression score was significantly higher than that from PRS related to response to stress in women, but not in men. In men, genetic contributions in depression score were not different across PRS. We did not detect significant differences in the estimated effects of PRS across sexes. However, the variance of depression score explained by PRS_{MDD} alone was 1.46 times higher in women than in men, whereas also considering the contributions from genetic responses to stress and SLE effects such difference in the variance of depression score was reduced to 1.16 times higher in women than in men. This may help to explain reports that the effect of sex on the onset of major depression is greater when psychological stress is not present; whereas for neuroticism, the personality trait from which the proxy for stress-sensitivity was derived, the impact increases with increasing adversity[79]. Notably, a significant PRS_{SS} prediction in men arose from a much smaller discovery sample size (GWIS N = 7,834 cases and 15,258) compared to the PGC-MDD GWAS discovery sample size used to weight PRS_{MDD} (PGC-MDD2 GWAS N = 50,455 cases and 105,411

controls)[150] or the GWEIS (N = 99,057 with a continuous measure of depressive symptoms).

Using the full sample, we detected nominally significant SLExPRS effects in *diathesis-stress* models for depression score only with $SLExPRS_{MDD}$ (R^2 = 0.08%. p = 0.044; as previously reported[431]) when using weightings from the largest PGC-MDD GWAS (49,524 cases and 110,074 controls). Such GxE effect remained undetectable using past PGC-MDD1 GWAS with a lower discovery sample size (9,240 cases and 9,519 controls; R^2 = 0.03%. p = 0.201). These results concur with the output from a previous test of the *diathesis-stress* model for depression reporting that whereas PRS weighted by PGC-MDD1 GWAS barely nominally predicted depression (p = 0.018), and therefore, being underpowered to test for a PRSxE effect, PRS weighted by an updated and largest PGC-MDD GWAS strongly predicted depression (p = 4.3 x 10^{-8}) and displayed a nominally significant *diathesis-stress* interaction in their full sample (p = 7.6 x 10^{-3}) and in women (p = 2.2 x 10^{-3}). Our results suggest that the interaction effect of PRS_{GxE} is smaller than PRS_{SS}, of whose effect in turn is smaller than PRS_{MDD}. Conversely, the sample size of the GWIS (N = 7,834 cases and 15,258) is lower than the sample size of the GWEIS (N = 99,057) and much lower than the sample size of the PGC GWAS (N = 50,455 cases and 105,411 controls). Our findings suggest that such studies are still underpowered to detect significant $SLExPRS_{SS}$ and $SLExPRS_{GxE}$ interactions in a *diathesis-stress* framework. Therefore, much larger sample sizes are required to detect robust effects, particularly with $SLExPRS_{GxE}$. At current available discovery sample sizes, PRS_{MDD} is the best instrument to test the presence of a *diathesis-stress* interaction effect. However, robust $SLExPRS_{SS}$ interaction effects may improve prediction of depression symptoms further as sample sizes increase.

Studies like this remain limited by statistical power. The statistical power required to detect GxE effects is higher than for marginal main effects due to their weaker effect sizes. Hence, such studies require larger sample sizes to detect similar significance levels[482-484]. However, in addition to increasing

sample size, improving the accuracy and robustness of measures of both SLE and trait may provide a more powerful strategy to detect robust GxE effects[485] by reducing measurement errors. The availability of high-quality environmental data and larger sample sizes to achieve statistical power are the main challenges GxE studies must face[259]. Approaches such as the GWIS that do not require direct measures of the environment (e.g. SLE experienced) and take advantage of a much more abundant measure in current population-based cohorts (e.g. neuroticism score) to model a proxy trait (e.g. stress-sensitivity) could help to address such limits. Considering the GWIS sample size (N = 7,834 cases and 15,258) and the significant prediction of PRS_{SS} detected, our findings enhance the potential of the GWIS approach to complement other tools in psychiatric genetic research[466] and maximize the power to detect relevant genetic factors underlying the aetiology of depression, particularly in men. Conducting genome-wide approaches split by sex may help to improve the accuracy of sex-specific effects; although a greater gain would likely arise from maximizing sample sizes[154].

By stratifying the *diathesis-stress* model by sex, our aim was to investigate GxE effects as a potential difference between women and men in genetic stress responses. Female sex increases liability to and shows higher prevalence of depression[48]. Differences in GxE effects displayed between girls and boys have been reported for adolescent depression[486,487]. We detected a significant $SLExPRS_{SS}$ interaction surviving multiple testing correction in men (p = 2.1 x 10^{-3}; PRS_{SS} weighted by GWIS of a proxy for stress-sensitivity at p-threshold = 0.01), which explained an extra 0.42% of variation in depression score in addition to the 0.50% explained by additive main effects on MDD. Compared to women, the estimated GxE effect size of such $SLExPRS_{SS}$ interaction in men was significantly higher (p = 1.06 x 10^{-2}), supported by a significant $SLExPRS_{SS}xsex$ interaction in the full sample (p = 1.00 x 10^{-3}). This suggests an additional risk of depression in men with high inherent stress sensitivity, not captured by main additive contributions to MDD. Consistently, the analysis by categories of SLE reported for this

$SLExPRS_{SS}$ interaction showed that, in men who reported high number of SLE (3 or more), the depression score increased with increasing PRS_{SS} score (p = 1.84 x 10^{-4}; **Table 6.3**). This effect was also nominally detected in the high SLE subset in the full sample (p = 0.047) but not in women (p = 0.71). To extend this, we assessed the *diathesis* effect of PRS_{SS} at the p-threshold displaying the strongest interaction in the full cohort (p-threshold = 1 x 10^{-5}). Although such $SLExPRS_{SS}$ interaction effect was not significant (p = 0.13) at this alternative GWIS threshold in the continuous model, we detected a nominally significant effect of this PRS_{SS} in those participants reporting a high number of SLE (N = 775, p = 7.37 x 10^{-3}), and in women (N = 490, p = 0.03), but not in men (N = 285, p = 0.11), perhaps due to the low sample size combined to the fact that the latter PRS_{SS} tested captures the aggregated effect of just 5 independent top SNPs from the GWIS of a proxy for stress-sensitivity (p < 1 x 10^{-5}). Consistent with the *diathesis-stress* theory[225], the increase in depression score seen in those individuals, particularly in men, with high PRS_{SS} and reporting high number of SLE suggests an additional risk of depression due to the influences from a genetically driven stress-sensitive trait. Although more statistical power is required to draw robust conclusions, our results support patient's sex as a relevant factor of stress-regulation in stress-related disorder. Differences in the aetiology of depression between sexes have been already suggested by previous tests of the *diathesis-stress* model[169,431], including in this sample, reporting differences in $SLExPRS_{MDD}$ interaction effects nominally significant between women and men[169,431].

In summary, alternative genome-wide approaches show the potential to identify relevant genetic variants in liability to depression that differs from those contributing to the additive main effects of MDD detected in standard GWAS. Overall, the evidence suggests a lack of power to robustly detect and replicate genetic variants with an effect in MDD as a response to environmental stress, rather than the absence of such variants. Consistent with results in twin-studies of recurrent depressive disorder reporting a departure from a model of additive genetic contributions[82], we showed that contributions from both PRS_{SS} and PRS_{GxE} are relevant in depression score

and not accounted for by PRS_{MDD} or SLE contributions. The detection of GxE effects using variants displaying modulatory effects of psychological stress with potential negligible main effects (i.e. not detected on GWAS for depression) has implication for future research strategies. Although replication and larger samples are required to find robust evidence, we show the potential benefits of identifying genetically mediated and modulator effects derived from these approaches. Understanding how men and women differentially respond to stress and identifying different stress-response mechanism underlying depression could help individuals to benefit from different types of prevention for stress-related disease.

Chapter 7 Summary and general discussion

7.1 Summary of main findings

The main aim of this thesis has been to advance our understanding of the genetic response to psychological stress underlying MDD. Throughout the last decades, robust evidence has arisen for an effect of both genetic and environmental risk factors such as SLE on risk of MDD. There is currently growing interest in understanding if, when, and how genetics and environmental risk factors interact, and the effects of such interactions on the development of MDD and other psychiatric disorders. Although many aspects surrounding the complexity of GxE remain unclear, I hope that the studies presented here contribute and help to push forward the research on this field.

In **chapter 1**, I presented the background of this thesis, providing the knowledge, evidence and concepts required to understand where we are in understanding how genetic effects responds to environmental exposures on liability to MDD and to show from where, how and why this thesis comes from.

In **chapter 2** and **chapter 3**, the main focus was a new trait conceptualized as a proxy for sensitivity to stressful and adverse life events dependent on MDD. In **chapter 2**, I conceptualized such a proxy as the higher levels in neuroticism shown in individuals with MDD. I detected putative loci involved in such trait through a genome-wide interaction study identifying a potential link with alcohol-related traits, and a significant gene-based association with *ZNF366*. I showed that the MDD-dependent stress-sensitivity component significantly improves the prediction of MDD, independently of main additive contributions from loci associated with both neuroticism and MDD.

In **chapter 3** I detected that genetic contributions to the MDD-dependent stress-sensitivity proxy were enriched within sets of genes responding to

glucocorticoid signalling, showing that variation in MDD risk within glucocorticoid signalling pathways was higher than expected, particularly due to genetic contributions to the proxy for stress-sensitivity. The findings pointed to the disruption of circadian rhythms as a potential mechanism for stress to lead to MDD and stress-related disorders.

In **chapter 4**, **chapter 5** and **chapter 6**, the focus was shifted to test quantitative measures of depressive symptoms, rather than dichotomized MDD diagnostic status, and to incorporate reported measures of recent SLE in order to investigate GxE effects underlying depression.

In **chapter 4**, following recent evidence supporting the *diathesis-stress* model for depression, I detected a significant GxE effect in liability to depression. Individuals with an inherent genetic predisposition to MDD that reported high number of recent SLE were at higher risk of depression. Therefore, my results, albeit weak, were consistent with previous evidence to support the *diathesis-stress* theory. These findings, supporting the presence of a GxE effect underlying MDD at an individual's genetic vulnerability level, validate the implementation of genome-wide GxE studies at the SNP level such as the one presented in **chapter 5**.

In **chapter 5**, I assessed GxE between SNPs and reported SLE, aiming to identify genetic variants associated with depression through genetic response to recent environmental stress. Applying gene-based tests into GWAS summary results, I first identified genes directly associated with depressive symptoms. Afterward, I reported two SNPs with genome-wide significant GxE effects in *PIWIL4* and *ZCCHC2,* and a SNP with genome-wide significant association with the combined main additive and GxE effect upstream *CYLC2*. In gene-based tests analyses, *MTNR1B* and *PHF2* were also significantly associated with GxE and the joint effect, respectively. *Dependent* SLE were shown to be heritable, but not *independent* SLE, supporting the hypothesis that *dependent* SLE are likely reliant on an individual's own behaviour, and are also likely to be strongly driven by genetic variability. PRS aggregating the GxE effects improved prediction of

depressive symptoms beyond PRS weighted by main additive effects of MDD. Furthermore, I showed that the aggregated GxE effect overlap with a wide range of stress-related conditions providing evidence that an underlying genetic architecture of stress response is shared between depression and common stress-related comorbidities. Whereas significant GxE effects came primarily from the analyses of GxE using *dependent* SLE, significant PRS prediction of stress-related traits came primarily from derived weights of GxE using *independent* SLE.

In **chapter 6**, I brought together data from **chapter 2**, **chapter 4** and **chapter 5** to investigate the relative contribution to liability to depression of stress-sensitivity and stress-response effects estimated in **chapter 2** and **chapter 5**, respectively. I applied these weightings within the *diathesis-stress* framework to test whether my new approach using weightings for genetic responses to stress in order to test for a GxE effect underlying MDD, rather than main additive effects of MDD, could improve the results from the standard *diathesis-stress* model implemented in **chapter 4**. I showed that both reported SLE, genetic sensitivity to stress, genetic response to reported SLE and genetic additive effects of MDD significantly contribute to increase risk of depression, both independently and together. At current sample sizes, stress sensitivity and stress response effects could not improve the *diathesis-stress* model implemented in **chapter 4,** but my results suggest that alternative tests such as the ones implemented in **chapter 6** could provide better predictions of GxE effects underlying MDD given larger sample sizes.

7.2 Discussion of thesis findings

7.2.1 Disentangling the genetic complexity of MDD

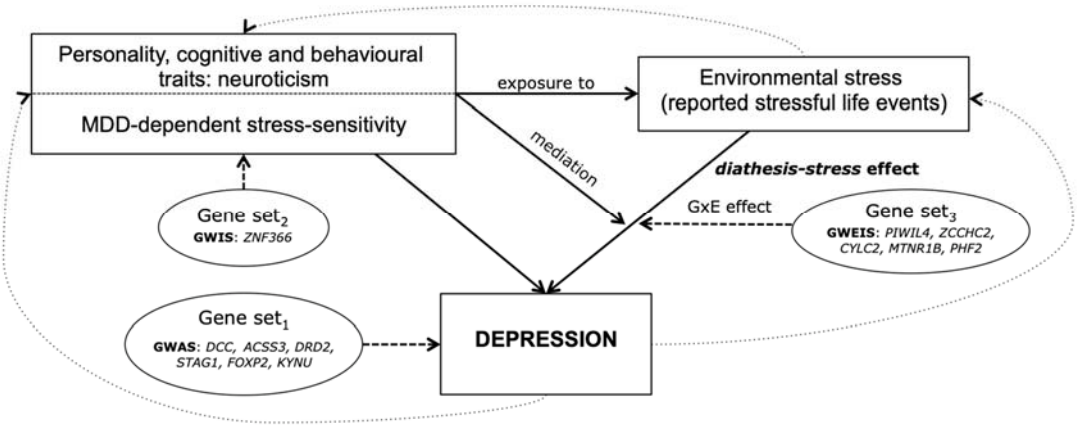


Figure 7.1 Schematic representation of aetiological mechanisms underlying MDD covered in this thesis. *Gene set$_1$ represent genes with a direct effect on MDD that could be identified by genome-wide association studies (GWAS). Gene set$_2$ represent a set of genes with an effect on liability to MDD through genetically mediated phenotypic traits with a direct effect on MDD; alternative analysis such as the genome-wide interaction studies (GWIS) could identify such genes. Gene set$_3$ represent a set of genes with modulatory effect on the direct effect of stress on MDD. Such genes could be identified through genome-wide by environment interaction studies (GWEIS). Arrows show some of the interrelations between risk factors and depression addressed in this thesis that reflect the complex system biology underlying MDD. Figure adapted from Gonda et al.*[488]

The aetiology of MDD is highly complex and heterogeneous. It depends on many genetic and environmental factors. Therefore, and as Kendler and Eaves already stated more than 30 years ago, to fully understand the complexity of the disorder, we need to understand the complex interplay

between risk factors. As seen above, there are many interrelations between genetic and non-genetic factors that may shape the aetiology of MDD. In **Figure 7.1**, I outline the interrelations underlying MDD examined in this thesis. Loci and genes (gene set$_1$), such as the ones identified in **chapter 5** using a gene-based test of summary statistics from the UK Biobank GWAS for a depression score, are likely to have a direct effect on liability. When I started my PhD in early 2015, no GWAS had detected loci associated with MDD. The largest PGC study, at that time, had been conducted in a sample including 9,240 cases of MDD and a similar number of healthy controls and had failed to detect any associated locus[147]. However, soon the first successful studies were reported and four years later the latest and largest meta-GWAS of MDD reported 269 genes associated with depression, involving 807,553 individuals including 246,363 cases[151]. Notably, the first study to detect significant loci achieved that by shifting from the mainstream approach, which was (and still is) based on increasing sample sizes while assuming that the inclusion of enough individuals will balance any phenotypic heterogeneity, to a specific study design focused on a sample of only female cases with a severe clinical depression phenotype with the same methodology and genotype platform applied to the entire cohort[149]. Together, recent successful studies underscored the fact that alleles conferring a higher risk of depression leading to, eventually, the manifestation of the clinical disorder can be found; we require enough power and the proper study designs to detect them.

Personality, cognition and behaviour may shape both the aetiology of illness and the environment. Several genetic factors contribute to such heritable traits (gene set$_2$). Thus, several genetic factors underpinning these inherent traits likely also contribute to mediate the effects of environmental stress in the pathoetiology of MDD. Potentially, through a wide range of mechanisms (e.g. exposure to stress, sensitivity to stress, perception of stress, regulation of pathways involved in stress-response systems, such as glucocorticoid-regulated systems and circadian rhythms, etc.) and thus, indirectly contribute to liability to MDD, and other stress-related conditions. In **chapter 2**, I

introduced a potential approach to partially target this genetic component by modelling of a proxy trait for an inherent MDD-dependent sensitivity to stress. Genetic contributions to this inherent trait (gene set_2) were shown to increase the risk of MDD beyond contributions attributed to genes with a direct effect on MDD (gene set_1). As shown in **chapter 3**, such contributions were enriched within sets of genes selected as candidate respondents of glucocorticoid receptor signalling. In addition to the foregoing, there are other genetic factors (gene set_3) that modulate the effects of environmental exposures. The added risk of depression due to GxE effects was validated in **chapter 4** and **chapter 6** supporting the *diathesis-stress* theory. Furthermore, several genes displaying GxE effects were detected in **chapter 5** through GWEIS. In **chapter 6,** I showed that contributions from all set of genes represented in **Figure 7.1** were significantly relevant on risk of depression. This high heterogeneity of potential mechanism reflects the high phenotypic heterogeneity seen in MDD, in which different causal pathways and distinct genetic and physiological mechanisms may underpin different subtypes of MDD patients. Hence, a diverse range of approaches to integrate different aspects of the pathophysiological system underlying MDD, including among others the ones presented in this thesis, should help to detect new genetic risk factors. These genetic factors may contribute to stratify MDD based on genetic and pathophysiological mechanisms underpinning, at least, subtypes with strong genetic components related to an individual's response to psychological stress. These, perhaps, also contribute to the shift from the traditional dichotomized nosology of MDD to new ways of classifying MDD patients using alternative transnosology approaches such as the Research Domain Criteria (RDoC)[489,490]. Consistent with this, I have introduced two alternative hypothesis-free genome-wide approaches to the standard GWAS: GWIS and GWEIS. Genes with direct effect on MDD detected on GWAS may be unfeasible as therapeutic targets, as they are likely involved in basic functions[488]. However, alternative approaches may reveal new risk variants, which may not contribute directly to the disorder, but have relevant contributions to liability representing thus a subset of potential therapeutic

targets[488]. Incorporating environmental data in genomic analyses such as the ones presented in this thesis may boost the detection of such therapeutic targets, not only in MDD, but in other fields as well.

Currently, available large population-based cohorts containing data on genetic factors, medical conditions and measures of SLE, allow us to study the interaction between genetics and SLE underlying the aetiology of MDD using different strategies. These strategies should allow a better understanding of the effects of SLE on liability to MDD, provide insights on the biological mechanism underpinning genetic sensitivity to stress and provide insights on the biological mechanism underlying the link between stress and MDD. This knowledge may provide potential applications and implications for policy and treatment, among others, to the health system and medical research.

7.2.2 The proxy for stress-sensitivity as a candidate endophenotype for depression

In **chapter 1**, I introduced the endophenotype concept, sometimes called "internal phenotype", as an entity employed to detect causal mechanisms between the genetic aetiology underlying a psychiatric disorder and the final manifestation of its clinical symptoms. Personality traits are thought to perform such role, neuroticism being the best candidate for MDD reported to date[198]. However, the benefits of neuroticism as endophenotype may be overestimated, as it is also a highly complex trait that does not seem to be more genetically simple than MDD and is placed between the causal pathways and the clinical diagnosis of MDD, but at a similar clinical phenotypic level. Nevertheless, the MDD-dependent stress-sensitivity trait conceptualized in **chapter 2** could be a better candidate. As this trait is dependent on MDD and neuroticism levels, it may involve fewer genetic factors than either MDD or neuroticism itself, involving fewer pathways and representing a closer trait to a genetic level in the causal chain to clinical disorders. Although it has not been tested, such proxy for stress-sensitivity as a trait is likely to meet the criteria to have the credentials of an

endophenotype of MDD. It is likely that it also meets the criteria to be an endophenotype of neuroticism, as I reported that: the proxy for stress-sensitivity is an inheritable trait; it predicts and associates with MDD status; although the detection of its genetic component is dependent on MDD, as this inherent trait reflects changes in neuroticism levels (and neuroticism is state-independent of MDD), it must be present regardless of the manifestation of MDD; as it is dependent on MDD and inheritable, it is likely to co-segregate within families; and to be enriched in unaffected family members. Furthermore, as I showed in **chapter 3**, its genetic component is enriched in sets of genes linked to glucocorticoid receptor binding sites likely triggered by cortisol signalling through glucocorticoid response elements. Morning cortisol and cortisol awakening response have been reported as potential endophenotypes for depression[192]. Therefore, the proxy for stress-sensitivity, as endophenotype, may improve the power to identify genetic risk factors contributing to liability, help to better understand some pathophysiological pathway involved, and enhance stratification of MDD.

7.2.3 The proxy for stress-sensitivity as mediator of stress response

The genetic effect of such an “endophenotype” may mediate the association between stress and MDD, but also the association between stress and other conditions linked to both sensitivity to stress and MDD. This could be the case of alcohol-related phenotypes (discussed in the next section) and other personality, cognitive and behavioural traits. Perhaps via regulation/disruption of circadian rhythms, thus implicating the circadian clock in such mediation between association of stress and disease as suggested in **chapter 3.** In **chapter 3**, I reported higher genetic variability in MDD risk conferred by genetic contributions to the stress-sensitivity proxy within glucocorticoid-related pathways, consistent with the function of *ZNF366,* the single genome-wide significant gene detected from the GWIS in **chapter 2**. *ZNF366* encodes for DC-SCRIPT, a zinc finger family protein preferentially expressed in dendritic cells involved in repressive activity on transcription through interaction with CtBP1[491], a C-terminal tail-binding protein that acts

as repressor of multiple transcriptional factors and is integrated in glucocorticoid receptor complexes[492]. Thus, *ZNF366* regulates gene expression mediated by glucocorticoid receptor[331]. *ZNF366* is the epitome for the genetic architecture modelled in **chapter 2,** supporting such genetic architecture as a truly underlying genetic component to a stress-sensitivity proxy. Furthermore, *ZNF366* specificity to be expressed in dendritic cells supports the need highlighted in **chapter 3** of targeting investigations on genetic response to glucocorticoid signalling as response to environmental stress in neuronal regions. Overall, the stress-sensitivity proxy may reflect the genetic architecture of sensitivity to the transcriptional effects triggered by cortisol signalling after experiencing stressful and adverse events, which Arloth *et al.* suggested may mediate the risk of MDD[377]. Notably, it has been suggested that genetic variation in cortisol-related neuroendocrine pathways may mediate the controversial interaction between polymorphisms in the *5-HTT* gene and stress (including childhood maltreatment and broad life stress) in current or lifetime depression, which may explain the lack of replicability and robust evidence supporting such an interaction[493]. Consistent with a system model with a high number of complex interrelations as represented in **Figure 7.1**. Perhaps, the stress-sensitivity proxy could contribute to disentangling such a puzzle.

7.2.4 The relevance of stress-sensitivity on treatment of comorbid alcohol dependency

As introduced in **chapter 3**, dysfunction of the HPA axis and glucocorticoid-response pathways is associated with alcohol dependence and problematic related behaviours, but also craving and damage in the brain's reward system[362]. Interactions between such alterations in stress-response systems and stressful exposures also increase the liability to alcohol-related disorders[362,363]. Thus, the vulnerability to alcohol abuse likely arises from the interplay between genetics and the environment, including GxE effects[494]. In regards to the proxy for stress-sensitivity, the strongest association detected in **chapter 2** through GWIS was in the *PTP4A1-PHF3-EYS* locus. This locus has been associated with several alcohol-related phenotypes[326-330], and in

alcoholic individuals the abuse of alcohol may up-regulate *PHF3* expression in the frontal cortex[495]. In **chapter 6**, I showed that the MDD risk (*diathesis*) conferred by the aggregated effect of only 5 independent SNPs from suggestive loci for the proxy for stress-sensitivity ($p < 1 \times 10^{-5}$) increased the risk of depression in those individuals who reported high number of recent SLE. 3 of these SNPs were genotyped from intronic regions of *OPCML*, *PHF3* and *EYS*. In addition to *PHF3* and *EYS* (that form part of the *PTP4A1-PHF3-EYS* locus), *OPCML* is also associated with alcohol dependence, as well as with depression, schizophrenia and behavioural disinhibition[496-498]. This gene encodes a protein required for coupling between G proteins and opioid receptors thus being essential in opioid signalling. Endogenous opioid systems are linked to the development of alcoholism[499]. Moreover, findings in **chapter 6** suggest a potential role of PRS weighted by the stress-sensitivity effect as moderator of the association between SLE and depression, particularly in men. Curiously, the variant in *EYS* (from the *PTP4A1-PHF3-EYS* locus), a gene also detected as genome-wide significant in the latest meta-analysis of MDD[151], is a potential eQTL for *LGSN,* the most significant gene in males in a study of sex differences for MDD[334].

The link between sensitivity to stress and depression is consistent with a potential relationship between depression, SLE and coping strategies reported in **chapter 5**. This suggests that genetic responses to environmental stress may modulate adaptive behaviours such as smoking or alcohol consumption, among others. Differences between women and men on genetic stress responses (further discussed in **chapter 4** and **chapter 6**) could help explain differences between sexes reported in the literature, not only in depression, but also in other stress-related conditions. In accordance with this theory, sex-specific differences in developmental pathways for depression have been suggested proposing a specific gene-environment interplay leading to alcoholism in men[53]. Significant three-way interaction between sex, depressive symptoms and drinking strategies to alleviate stress have been published, showing that the motives to drink alcohol as a coping strategy were more strongly associated with problems among men reporting

'higher' in depressive symptoms than among women[500]. Unlike women, men are at greater risk for increased drinking and also tend to cope with stress by consuming more alcohol[52].

These findings are consistent with previous evidence for a link between environmental stress (including genetic responses to), MDD and alcohol-related conditions[335-340]. These could help to detect genetic factors with a mediator role on the effects of stress, behavioural traits and depression (i.e. a mediation between the interrelations of all three "square boxes" in **Figure 7.1**), maximize the power required to identify relevant genetic factors in stress-response systems underlying not only the aetiology of MDD but also the aetiology of other stress-related conditions such as alcohol-related traits, and thus lead to treatable pathophysiological mechanisms that may inform better personalized treatment for comorbidities such as alcohol dependence and for at least a subtype of MDD patients.

7.2.5 Sex-specific differences in genetic responses to environmental stress

The results detailed above add to the evidence introduced in **chapter 1** to suggest that sex-specific differences may be present in genetic response to environmental stress and adversity. Among patients with panic disorder, sex differences have been reported in the types of SLE experienced, coping styles, agoraphobia and physical functioning, suggesting that sex-specific intervention could be required to provide better assessment and treatment to patients with panic disorders[501]. As was concluded in **chapter 6**, sex may be an essential factor to consider in genetic responses and the regulation of stress response systems in stress-related disorders. However, to draw robust conclusion we require stronger evidence from more independent studies. Understanding such differences in genetic responses to stress between women and men underlying depression may help to elucidate differences seen between both sexes in several aspects of the illness such as disease prevalence, symptoms reported or copying strategies[49-53]. This may be

extended to other stress-related conditions that present sex-specific differences.

7.2.6 Genetic-response to stress as a single trait within pathogenesis of stress-related disorders

Many human disorders, both mental and physical, could be categorized as stress-related disorders. Most of them are heritable and many of them co-occur and overlap in time, being comorbidities of each other. As suggested in **chapter 5**, genetic and physiological mechanisms underlying stress response systems are likely to be shared and partially explain pleiotropic effects detected between some of these disorders[502]. Although the genetics of stress response likely plays a central role in the pathogenesis of many stress-related disorders, it may have its own polygenic architecture that constitutes a unique entity (i.e. the genetic architecture of stress response systems). However, to investigate genetic factors integrated in stress response pathways that increase liability to illness, we generally require of a target phenotype to assess its detrimental effects. Correlation of SLE is stronger with psychiatric disorders than with medical or physical illness[444]. Furthermore, MDD is one of the disorders more closely linked to such pathophysiological mechanisms, together with anxiety and posttraumatic stress disorder, although contributions of genetics of stress extends to other psychiatric disorders as well[92]. Therefore, to model genetic responses and GxE effects, through GWEIS or other alternative strategies, on depressive symptoms and MDD may be a promising initial approach to investigate shared aetiologies with other illnesses through stress-response systems. In **chapter 5**, I provided evidence of a partial shared genetic overlap between genetic response in depression and other mental and physical health conditions linked to the negative effects of stress such as schizotypal personality disorder or cardiovascular diseases, reflecting a potential pleiotropic effect and shared aetiology due to the genetics of stress response. Genetic responses to stress may shape personality and adaptive behaviours that will in turn mediate the association between genetic factors underlying stress response pathways and stress-related traits. As the size of

discovery samples to estimate genome-wide GxE effects increase, it is likely that new associations driven by genetic responses to stress across stress-related traits arise.

7.2.7 PHF, a family of stress response genes

Perhaps one of the most relevant findings from this thesis, with regards to identifying genetic factors involved in genetic responses to environmental stress, is the significant detection of *PHF2* using a joint effect in GWEIS reported in **chapter 5**. As Kraft et *al.* reported, joint test of marginal genetic effect and GxE has, in general, greater power than a simple marginal test (as implemented in GWAS) when the genetic effect is only expressed in individuals exposed to the environmental risk factor[233]. In line with this, Wong *et al.* used whole-genome screening of functional variants to perform a GWAS including 203 mild to moderate MDD cases and 193 controls while considering the levels of stress experienced by the control population. They used a sample of Mexican American population from Los Angeles comprised of individuals born in Mexico most of which were recent immigrants that had experienced significant hyperactivation of the HPA axis as consequence of suffering recent SLE related with challenge, distress and acculturation issues[343]. They detected common and rare functional variants associated in 44 genes, but only associations within *PHF2F1B*, also known as *PHF4*, gene replicated in a European cohort. Furthermore, they found that *Phf21b* hippocampal gene expression is significantly decreased in rats resilient to chronic restraint stress when compared with non-chronically stressed rats and thus modulates the chronic stress response. Therefore, they reported *PHF21B* as a gene associated with MDD and stress response[343]. In mouse, the loss of *phf8* effects resilience to anxiety and depression-like behaviours, both with high contributions from stressful factors[344]. *PHF8* (also known as *KDM7B*) is a paralog of *PHF2* (also known as *KDM7C*) with antagonistic effects that compete for binding ribosomal DNA. While *PHF8* promotes ribosomal RNA gene transcription, *PHF2* plays a repressive role[459]. Both genes belong to a subfamily of Jumonji domain-containing histone-lysine demethylases (Jmj-KDMs or KDM7s). As Wong *et al.* highlighted, *PHF21A,* a

prominent gene from the same family, is highly expressed in the brain, encodes a component of a histone deacetylase complex and inhibits transcriptional activity during neurodevelopment periods[503-505]. *PHF2,* like *PHF3* discussed previously in regard to the stress-sensitivity proxy, *PHF21B* (also called *PHF4*), *PHF21A* and *PHF8*, encodes for a PHD finger motif protein (PHF). PHF proteins specifically bind a wide range of histone marks associated with the regulation of transcriptional activity and enriched at gene promoters[459,506,507]. This is a family of sophisticated histone code readers that interact with different histone modifications in order to regulate gene expression through the recruitment of transcription factors and protein regulators. Thus, interpreting and interplaying with the epigenome[508]. Furthermore, they have been related to oncogenesis and epigenetic inflammatory signalling[506]. Noteworthy, PHD finger motifs, and in particular PHF proteins, have been suggested as candidate targets with a strong therapeutic potential[507,509]. PHF may play a key role modulating stress-response through epigenetic regulation and chromatin remodelling[510], thus providing directions for future research on genetics of stress response.

7.2.8 Relevance in a clinical setting

The findings presented in this thesis do not have direct and immediate implications for clinical practice; the importance of such findings is yet of limited clinical relevance. Nowadays, MDD diagnosis relies predominantly in the patient's behaviour and symptoms rather than on clinical tests for biomarkers. In this sense, the findings presented in this thesis contribute to identifying new loci with potential as biomarkers for MDD risk and to help to advance in our understanding of key pathways involved in pathogenesis and stress-response. However, the extremely high heterogeneity and complexity of MDD limit the immediate utility of findings from genome-wide studies. Thus, their potential use is still far from clinical application. Nevertheless, a better understanding of the genetic mechanism and risk factors leading to MDD could support more efficient risk profiling, diagnosis, tailoring therapy and prognosis. Overall, the findings presented represents part of the vast

literature and accumulated knowledge required to, in the future, drive more effective prevention and treatment strategies.

We just recently started to identify genetic risk variants for MDD over the last few years. The identification of new risk loci opens broad avenues for future mechanistic studies that may contribute to future clinical relevance. In this sense, results obtained from present and future genome-wide studies (e.g. GWAS, GWIS, GWEIS) may contribute to ground future clinical trials. There are a few examples of translation of genetic findings from GWAS of complex traits towards clinical applications[511]. The challenge now is turning this knowledge of the genetic and biological mechanisms of MDD into clinical applications. However, before translating findings from genome-wide studies of MDD into clinical research some other challenge must be faced[512], such as confirming target gene predictions, understanding target gene functions and their role in MDD risk, and identifying prioritized targets for drug development, among others.

Any such individual common genetic variants detected in genome-wide studies is incapable of predicting or stratifying MDD alone. However, PRS have a greater precision of risk identification and could be relevant in clinical settings. The estimation of an individual's genetic risk would be one of the earliest measurable contributors to risk of MDD, with potential clinical utility on early disease detection, prevention and intervention. However, the utility of PRS is currently limited by its simplicity. Nowadays, clinical risk prediction usually relies on demographic characteristics, family history, health parameters, life factors and experienced events. Currently, PRS lack predictive power, but ever-increasing sample sizes contribute to increase their power with greater accuracy. Therefore, recent advances in PRS have major implications for its clinical potential. PRS have been already extensively applied over the last decade, not in a clinical setting for the prediction of individual's risk but in applications that facilitate new experimental designs and discoveries[511]. There is already evidence supporting the clinical utility of PRS in other complex diseases[513] and, in psychiatry, the possibility of adapting PRS for clinical use has been already

considered[514,515]. In other diseases, PRS have already showed promise in their capability to identify subgroups of patients that may benefit from more personalized strategies[513]. However, PRS are not expected to be a diagnostic tool able to classify between healthy and depressed individuals but likely to be used in prognosis and prediction of the likelihood that the onset of MDD will occur in each individual or subgroup of individuals. PRS could be used as a tool to determine a patient's position in a distribution of genetic risk so that patients falling on the tail above a threshold would be consider of high risk[514], allowing us to identify groups of individuals who may benefit from the knowledge derived from genome-wide studies on their probabilistic susceptibility to MDD or to respond adversely to environmental stress. In combination with other clinical risk factors, distinct PRS could support stratification of patients with distinct degrees of absolute risk and thus, could help drive clinical decision-making. Hence, a plausible implementation of PRS into a clinical setting in the near future is the identification of subsets of individuals with higher risk of MDD on the basis of genetic factors in combination with other clinical risk factors. As our knowledge of the genetic basis of MDD increases, PRS could support a more targeted therapeutic approach[516]. However, large-scale prospective studies assessing the clinical utility of PRS are still required. At the end of the day, we must acknowledge that for disorders and potential subtypes of MDD with low heritability, genetic scores may never become clinically useful. Moreover, there are still some scientific, clinical and social hurdles to bring PRS into practice, such as, among others, the uncertainty about causal variants determining the genetic risk estimated in an individual, the lack of knowledge about how to interpret PRS or their unwanted psychosocial impact[517]. Besides, the use of PRS in clinical contexts, most prominently in psychiatry, could have ethical implications, support reductive interpretations and feed into problematic assumptions regarding psychiatric disorders[516].

7.3 Methodological remarks and limitations on GxE research

There are still many methodological challenges to tackle in GxE research, including some limitations shared with other genome-wide studies. Unlike GWAS, when I started my PhD the first genome-wide GxE study in regard to MDD had yet to be conducted. Hence, we are still in the early stages of success and, in comparison to GWAS, the growing explosion of genome-wide GxE approaches is likely to be slower. Further progress in GxE research will require, at least, addressing the following methodological issues and limitations.

7.3.1 Sample size, statistical power and false positive findings

Like other genome-wide approaches, sample size is a key limiting factor in achieving statistical power. Having sufficient statistical power to detect true signals and avoid false positive or negative results is a major concern for genetic studies in general. In **chapter 1**, I introduced how after Caspi *et al.* studies on 2003 there was a huge proliferation of GxE studies in candidate genes. Most of these studies were conducted on samples ranging between 100 and 3,000 participants. Nevertheless, larger sample sizes are required in order to conduct the well-powered direct replication studies that are essential for GxE findings to gain credibility[443]. Studies with small sample sizes would only be well powered under the assumption that common variants with GxE effects explain a fairly large proportion of the trait variance[274]. Thus, some suggest that most (if not all) of the positive findings reported on such studies were likely false positives[251,443], as it is not clear how robust and replicable these findings are[518]. Furthermore, most genetic effects found so far in psychiatric disorders explain far less than 1% of phenotypic variability so we know that, on average, like main additive effects modelled by GWAS, most GxE effects must be very small. However, it is hard to know what GxE effect

sizes are reasonable to predict MDD. For example, one of the largest GxE effects ever reported was detected for an interaction between smoking and a polymorphism in *CHRNA5* on smoking-related diseases which was equivalent to only 0.5% of the variance[519]. Assuming that a GxE effect may explain 0.1% of the variability in depression-related traits, then sample sizes ~10,000 individuals would be required, if we reject a null hypothesis at the 5% level ($\alpha = 0.05$). If GxE effects explained 0.01% of the variability, ~100,000 individuals would be required to conduct well-powered studies. Nowadays, samples over a 1 million individuals are available to conduct GWAS for MDD, and although sample sizes big enough to also perform reasonable well-powered GxE studies at genome-wide scale are being recruited (which should solve most of the limitations of previous candidate gene studies), sample sizes available to conduct GWEIS are yet a few magnitudes behind GWAS due to lack of data on measures of environmental exposures. Moreover, as discussed in **chapter 6**, to detect GxE effects requires, in general, more statistical power than detection of additive effects and, therefore, GWEIS may require even larger sample sizes than GWAS in order to detect effects at similar significant levels[482-484,520,521].

If GxE studies are underpowered, we would expect a high false discovery rate with a consequent low replication rate. Indeed, another main concern is the possible inflation of false positive findings when heteroscedasticity is not taken into account. However, as it was reflected by genomic inflation estimates (λ_{1000}) in QQ plots from GWEIS reported in **chapter 4** (**Appendix C**), this concern seems to be well solved with the implementation of a nonlinear statistical approach that uses robust estimates of standard errors[272]. Other tools, different to the one I implemented, also apply robust standard errors and have been also used in order to correct for such biases in GWEIS[113]. Which tool performs better in such analyses is still unclear and the choice of one above the other should probably depend on study design. Nevertheless, some of the most popular tools to conduct genomic analysis, like PLINK, allow testing for interactions but do not take into account such potential bias, resulting in an inflation of false positives in some scenarios,

unless you incorporate an external plugin specifically design to adjust for heteroscedasticity effects. Heteroscedasticity is a statistical phenomena unknown by most researchers, not only in psychiatric genetics but in other biomedical fields as well, that has been largely ignored by previous studies testing GxE effects. This could bias and lead to false positive findings like the non-inclusion of PCs to adjust by population structure which, in contrast, is a mainstream approach to properly adjust studies. Therefore, caution must be taken when selecting software to conduct GxE analyses.

As I showed in **chapter 4** and **chapter 6**, a promising alternative in order to detect GxE effects is to test interactions using PRS, rather than single SNPs, as they have much power to detect phenotypic variability and reduce multiple hypotheses testing. This approach is getting more popular. However, following the rationale behind using PRS to predict a trait weighting by the genetic effects on such trait, if we want to use PRS to predict GxE effects, logic should lead to think about using weightings for GxE or environmental-response effects. In **chapter 6**, I address this point and, although at current sample sizes this approach is not yet the best alternative to implement, the results suggest that, in the future, with good enough GxE estimates from large enough sample sizes, this could be the best practice. It must be said that potential limitations could arise in PRS approach when discovery and target samples come from different population ancestry. Therefore, it is important to target GWEIS not only to European-descent samples.

7.3.2 Improper implementation of control variables

Another consideration to consider is to properly control for the effect that covariates may have on GxE effects. This is something that most studies have also omitted in the past, as it was thought that confounding effects were already adjusted by fitting the proper covariates into linear models. However, Keller MC demonstrated that confounding effects on GxE effects could remain if covariate by gene and covariate by environment interactions were not also adjusted[424]. Let's take as example Caspi *et al.* findings reporting GxE effects between SLE and *5-HTT* gene on depression. Imagine that

females “hypothetically” were much more likely to report SLE (i.e. SLE is highly correlated with being female). Under such assumption, high SLE would be a proxy for being female and thus, Caspi’s results may reflect an interaction effect between *5-HTT* gene and sex. Therefore, the GxE effect reported would be attributable to an interaction between *5-HTT* gene and female sex, rather than SLE. The method proposed by Keller is as simple as adjusting the model by the interaction between *5-HTT* gene and sex. If then we still estimate a significant GxE effect with SLE, we could draw the robust conclusion that it is not confounded by sex. However, it must be considered that incorporating all covariate interactions into a model may overfit it and, although it should be acknowledged now that covariate interactions must be adjusted to properly adjust GxE models, it could be argued whether all covariate interactions must be fitted, rather than only those ones we consider that could have a confounding effect on the GxE. Perhaps, for example, interactions with PCs (if fitted into the model) could be discarded for the sake of avoiding overfitting.

7.3.3 Phenotypic and environmental measures: the quality of data

The instruments used to diagnose and measure both the phenotype and the environment are key factors, and its parameterization, including scale and units used, could have direct effects on the GxE modelling[425,430]. This is an important point to consider for those traits that do not have natural units. In fact, any non-linear transformation could create spurious GxE effects. This highlights the importance of being able to defend a particular choice of units (e.g. number of SLE, years of education, height in cm, standard deviations, etc.). Probably, the best approach on phenotypes without natural scale would be to standardize the units used to capture such a phenotype (e.g. score from self-reported depressive symptoms) in order to represent a trait that is distributable in your sample (preferentially under a normal distribution), so it can be measured by standard deviations from the mean and it is not arbitrary. In regards to traits like self-reported depression, a psychometrically defined phenotype to reflect the dimensions of complex self-reported traits,

rather than summation scores like the ones used in this thesis to assess depressive symptoms, may further improve GxE studies and the investigation of genetic associations with MDD[522].

7.3.4 The issue of self-reported data

As seen above, accurate phenotyping is a crucial factor for investigating GxE. Therefore, GxE research requires accurate and robust high-quality measures of the environmental exposure[261]. Studies with relative small sample sizes often use higher-precision prospective measures that reduce the odds of false positives, as they assess environmental variables with less measurement error than larger studies. Data on environmental factors such as SLE in studies with large sample sizes (if available) tend to use low-precision retrospective reports such as self-reported questionnaires. However, the primary method applied to measure SLE or other life events (e.g. childhood trauma, diet or physical activity) is through self-reported questionnaires. Self-reported questionnaires are an economical and quick method in research for assessing a wide range of constructs based on a battery of questions. The application of this kind of measures often carries a number of limitations. Furthermore, although some instruments are design to cover the same underlying effects or contextual environment, different questions may cover and emphasise different features of stress, the environment, or indeed of the illness.

In regard to this thesis, the most important issue lies with the temporal order of SLE reported and diagnosis recorded. To investigate GxE effects, we require SLE measures to cover a time period preceding MDD diagnostics or reported phenotypic data. It may not be feasible to collect new measures of SLE covering the time period preceding the phenotypic records available in population-cohorts. But if we now collect data on recent SLE, we also need to collect new phenotypic data on MDD status or depressive symptoms. This is a main limiting factor to have larger sample sizes available to conduct GxE studies. Another important issue concerning the use of self-reported questionnaires to measure the amount of stress that an individual has

suffered during a specific period of time lies with its validity and whether it can fully capture the construct of psychological stress. The amount of stress registered can vary across different events and different causes investigated to determine the state of such psychological stress (i.e. depending on the questions assessed) and the units used (e.g. number of stressful events suffered vs. severity of such events). These limitations including retrospective and response bias (i.e. information remembered and willing to provide, and preference towards giving particular answers) are extensive to the use of any self-reported questionnaire and can altogether affect the final data collected through the questionnaires[523]. Difference in instruments used may explain lack of replication. For example, the measures used in **chapter 5** for UKB participants, both for SLE and depression score, were less informative (i.e. covering a substantially reduced spectrum of symptoms over a larger time period) than for GS participants. Therefore, although UKB GWEIS was conducted in a substantially larger sample size, it may explain the lack of replication and significant GxE reported in UKB. Different findings reported in **chapter 5** between self-reported measures of SLE in GS (i.e. *total*, *dependent* and *independent* SLE) reflect how using different items that cover different environmental effects produce different results, highlighting the importance of using the same instruments to replicate results. It must be noted that despite its validity and robustness, such tools, like the List of Threatening Experiences used in this thesis, may not cover a wide range of minor events with potential long-term or mild to moderate contextual stress, that likely impacts on the final adversity faced by an individual and thus on liability.

7.3.5 The aetiological model for GxE underlying MDD

In addition to the foregoing, most instruments to measure environmental events are designed to cover only those events with potential negative effects, but omit the full range of events with potential positive contributions. Most etiological models investigated in psychiatric disorders to date have been conceptualized under the *diathesis-stress* theory. As shown in **chapter 4** and **chapter 6**, there is multiple evidence to support such theory in order to

understand the aetiology of MDD. However, such theory only covers one side of the full environmental spectrum to which we are exposed to. As introduced in **chapter 1**, there are a few alternatives to the *diathesis-stress* theory such as the *vantage sensitivity* theory, which only considers the positive contributions from the environment, or the *differential susceptibility* theory. Both the *diathesis-stress* and the *vantage sensitivity* theories have its own conceptual space inside the *differential susceptibility* theory (see **Figure 1.2**). Moreover, it is plausible to think that individual differences in sensitivity to environmental factors contributing to depression are not exclusive of negative exposures with adverse effects, but also of positive events with resilient effects. Perhaps, the same underlying mechanisms support both hypotheses. We would need uniform environmental information along the entire spectrum from positive to negative events to disentangle such question. In addition, if we consider a broader spectrum of contributions from the exposome, the odds of *crossover* interactions, rather than *fan-shaped* interactions, increase, supporting the *differential susceptibility* theory. Altogether, this would support the existence of alleles with "plastic" effects, rather than "risk" or "resilient" effects. In fact, in **chapter 4**, I suggested the possibility that some alleles could be operating as plastic alleles, rather than risk alleles. However, as noted, we do not know whether the same alleles are involved in both mechanisms (vulnerability and resilience). But, if the concept of genetic plasticity conferring sensitivity to both positive and negative environmental effects was true, it could explain the lack of replicability and the inconsistent findings reported by GxE studies in candidate genes. These studies may have tested polymorphisms that act as "plasticity alleles" while only considering one side of the environmental spectrum. Therefore, not only potential disadvantages, but also potential advantages, of genetic variability in sensitivity to environmental exposures underlying MDD should be considered. A better understanding of how genetic variants interact with the environment may lead to policies that provide the right environment to help patients reach their full potential[524]. Furthermore, understanding both desirable and undesirable effects due to the response to specific

environments as a function of an individual's genotype may lead to personalized and targeted policy interventions. Unfortunately, measures of positive life events are much more limited than measures of adverse life events in most cohorts with genomic data.

7.3.6 How should we interpret GxE?

In **chapter 1** I introduced a basic equation to study GxE as follows:

$$y_i = \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i x E_i + \epsilon_i$$

However, GxE effects have at least two possible biological interpretations that are indistinguishable through standard GxE approaches based on this model. These interpretations could be stated with the following equations (note that both are equivalent to the equation above):

$$y_i = \beta_0 + (\beta_1 + \beta_3 E_i) \cdot G_i + \beta_2 E_i + \epsilon_i$$

$$y_i = \beta_0 + \beta_1 G_i + (\beta_2 + \beta_3 G_i) \cdot E_i + \epsilon_i$$

On one side (1st equation), it would be the environment that moderates the association between the genotype and the phenotypic outcome, in which case the effect of the genotype depends on the environment. So, if both β_1 and β_3 are positive, the larger the effect of the environment, the larger the effect of the genotype on the final outcome. On the other side (2nd equation), another plausible interpretation could be that it is the genotype that moderates the association between the environment and the outcome. In this case, it is the coefficient on the environmental effect that depends on the genotype. Although the first scenario seems more plausible, from a biological perspective it is unlikely that either of these is the right interpretation alone, but rather both working together. Perhaps, in the future, more sophisticated approaches to study GxE effects identifying the underlying mechanism implicated allow disentangling such puzzle in order to decipher how GxE effects should be better interpreted. Mendelian randomisation approaches may help to identify causal paths and directions linked to the effects of SLE[525,526].

7.3.7 GxE vs. GxG effects

A limiting factor to interpret results from GxE studies is whether the reported estimates truly reflect GxE effects or, instead, they are due to gene-environment correlation[74,426]. As Kendler and Eaves introduced in their 3rd model, genetic contributions to exposure to SLE (or reporting of self-reported measures in retrospective studies) are likely to occur. Therefore, if we take again as example the interaction between *5-HTT* gene and SLE on MDD, even under the assumption that such interaction was robustly estimated and well-replicated in large enough sample sizes, we could not prove that it is the result of a pure GxE. The only way to tackle this is to ensure that the environmental measure does not capture any genetic influences or, if SLE does capture genetic influences, that GxE is not driven by gene-environment correlation. I addressed this in **chapter 4** and **chapter 5** by using SLE defined as *dependent* and *independent*. However, evidence supports either that both measures partially cover the same underlying effects (e.g. self-reporting bias), or the biological mechanisms involved in both the direct and GxE effects of each SLE measure differ, or both. It is important to prove that a detected significant GxE truly reflects an interaction between a genetic and environmental component if we want to properly understand the mechanism of how such association with MDD arise. Evidence suggests that to ensure that SLE does not capture genetic influences through depressive status, personality or behavioural traits, among others, is unfeasible using self-reported measures.

However, we should distinguish between what it is a real GxE and what derives from the 3rd model proposed by Kendler and Eaves about a genetic control of exposure to environment, or what others call “nature via nurture”. If an estimated GxE was driven by the genetic component and not by the environmental component of SLE, such estimate would arise from gene-environment correlation and would reflect a subtle GxG interaction rather than a GxE interaction. Although such findings could be relevant to liability for MDD. Such contribution of genetically mediated reporting of SLE may be driven by inherent traits such as sensitivity to environmental stress. As

previously discussed, identifying genetic contributions mediating the association between SLE and MDD could contribute to understand the aetiology of MDD, and epistasis could be another mechanism modulating the effect of environmental stress (i.e. GxGxE interaction). This has been supported, for example, by GxE studies on the risk of asthma as a result of smoking and polymorphisms in genes for xenobiotic metabolizing enzymes. It has been reported that epistatic GxG effects among such enzymes in smoking individuals increased the risk of asthma above the expected risk without an epistatic effect[435]. In fact, behaviour, coping style and personality traits influence susceptibility to stress, and their genetic contributions likely interact within genetics of responses to stress.

7.3.8 Early by late stressful life events and a life-course approach

There is ample evidence that interactions between SLE occur over lifespan (i.e. environment-environment interactions, ExE) and also contribute to the risk of MDD[527-531]. Furthermore, sex-specific effects may contribute to modulate ExE[532-534]. Three main theories have been developed in order to conceptualize a dose-response interaction between the number of SLE, or its severity, over the lifespan (see **Figure 7.2**) that could modulate liability to MDD: the stress-amplification, the stress-inoculation and the stress-sensitization theories[534,535].

Under the stress-amplification theory, the effect of earlier SLE (e.g. childhood trauma) enhances the effect of later life SLE on MDD liability[534]. Conversely, under the stress-inoculation theory, the effect of earlier SLE makes you more resilient to the adverse effects of later SLE, suggesting that mild to moderate levels of stress would boost the development of adaptive mechanisms contributing to better coping later in life and thus, reducing the adverse effect of later stress[536,537]. In fact, it has been suggested that not all SLE have negative effects[444]. Finally, under the stress-sensitization theory, individuals who experience SLE became more sensitive to the negative effects of later SLE and thus, their threshold to trigger symptoms of illness is reduced over

time[527,538]. This is consistent with the *diathesis-stress* theory, in which individuals are vulnerable to develop symptoms with low levels of adult psychological stress. Nevertheless, under the stress-sensitization theory, such individuals are vulnerable because they had experienced high levels of earlier stress, rather than because they carry a high load of risk alleles. This could contribute to the lack of positive findings to empirically support the *diathesis-stress* theory. It may also link to the kindling hypothesis for depression, in which sensitivity to SLE is increased by an episode of MDD, increasing the autonomy for the onset of new episodes of MDD[539]. Perhaps, all three forms of ExE co-exist. Hence, evidence for ExE and gene-environment-environment interactions (GxExE) along lifespan influencing psychiatric disorders suggests that genetic sensitivity towards the effects of stress (stress-sensitization) is dependent and modulated by previous cumulative interactions experienced over lifespan[479-481] and highlights the need to integrate SLE at different time-points in life[540]. GxE effects (particularly facing early childhood adversities) may result in neurological, physiological and/or cognitive-emotional consequences that contribute to modulate the effect of future GxE. This was supported by reported significant evidence in children with high PRS for genetic environmental sensitivity[480]. Those sensitive children exposed to negative environments exhibited increased sensitivity to SLE in adulthood. Conversely, those sensitive children exposed to positive environments early in life were significantly more resilient to adversity in adulthood through GxExE effects[479,481]. If MDD, as other psychiatric and stress-related disorders, was the result of a $GxE_1xE_2xE_3xE_4…xE_n$ interaction along lifespan, assessing a time-point interaction (e.g. $GxE_{n>4}$) in adulthood without taking into account modulating effects of previous environments, specially at early developmental stages, may result on negative or contradictory findings. Furthermore, as we have already discussed, GxE may be mediated by genetic contributions, such as an inherent genetic sensitivity to stress. Evidence points to a complex interactive system between the genome and the environment underlying the aetiology of MDD. Therefore, life-course GxE approaches including the full

range of environments from early childhood to adulthood may be required to elucidate stress-response mechanisms underpinning the development of depressive symptoms and to fully understand GxE effects. Furthermore, the presence of modern and exponentially increasing environmental inputs with potential stressful effects characteristic of a western lifestyle, which may not be captured and evaluated by current screening tools, may induce new GxE effects through environmentally-induced changes in effect sizes of causal alleles with genetic susceptibility effect to depression[541].

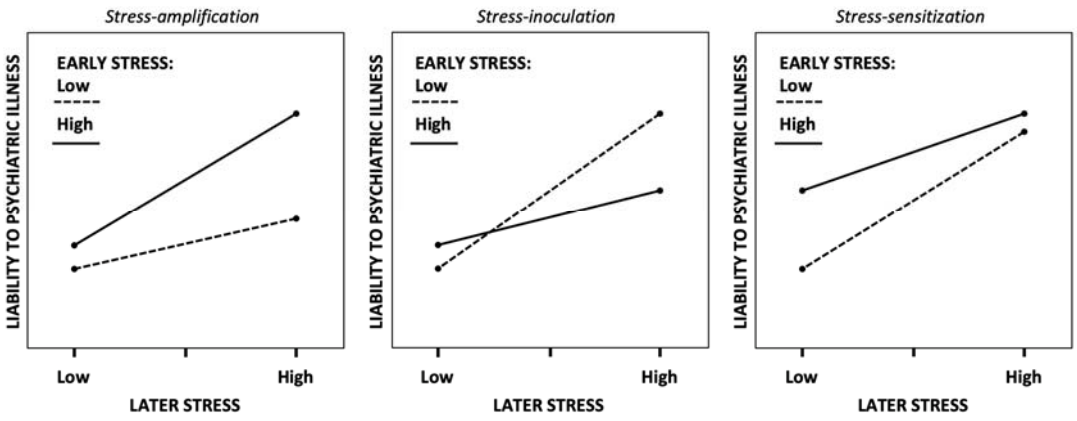


Figure 7.2 Schematic representation of early SLE by later SLE interaction theories. *Stress-amplification* (left): high adversity of early SLE increases risk of MDD in those individuals exposed to later SLE conferring high levels of psychological stress. *Stress-inoculation* (centre): high levels of stress in early SLE buffers against the adverse effects of later SLE. *Stress-sensitization* (right): high levels of early stress increase sensitivity to deleterious effects of later SLE, increasing the risk of MDD in those individuals experiencing low adversity in later SLE but with no differential risk in those experiencing high levels of later psychological stress. Figure adapted from Rudolph *et al*.[534]

7.4 Future perspectives

After a decade of scepticism about research on GxE in psychiatric genetics due to the inconsistency of findings, the dramatic increase in data and resources available foretell a promising future for research on GxE underlying MDD, as well as other psychiatric and mental illnesses. The identification of robust genetic risk variants from recent and future GWAS, GWIS and GWEIS, among other strategies, will provide the knowledge required to select new sets of candidate polymorphisms, genes and pathways with potential therapeutic effects. Although research on GxE may identify new variants with a very small overall effect on MDD, it could reveal new therapeutic pathways to target in therapeutic approaches. However, much remains to be investigated to fully understand the background of such disorder. Our actual knowledge of the complex interplay between genetic factors, environmental stress and stress response mechanisms underlying MDD is still limited. Furthermore, although MDD is a well-recognized public-health priority due to its huge economic and health burden growing over time, depressive disorders (and psychiatric disorders in general) are among the most underfunded illness in biomedical research[542]. This trend will need to reverse in the future. Hopefully, the work presented in this thesis contributes to demonstrating the relevance of GxE research, pushing forward this field, and attracting more funding from engaged institutions.

For the near future, I think progress can be expected from analysing the recent release of UK Biobank data from over 500K individuals with available data on a wide range of medical conditions and updated new mental health questionnaires covering a wide range of psychiatric symptoms (including neuroticism levels) and SLE. This data of great value provides a unique opportunity to complement the results I have presented by performing GWIS and GWEIS studies on substantially larger sample sizes, and thus to expand the line of investigation covered in this thesis. Meanwhile, though, better implementations of GxE genome-wide approaches in a time-efficient manner,

as discovery sample sizes increase, are a priority. The performance of GWEIS can be highly time consuming in sample sizes exceeding 100K individuals when robust estimates of standard errors are applied (i.e. when we implement a nonlinear statistical approach that uses Huber-White standard errors to correct for a potential inflation of the GxE effect estimated due to heteroscedasticity), the exposure has a wide range of levels, and we test for large numbers of genetic variants (~1M SNPs or more). Therefore, new computational strategies may be needed to conduct genome-wide GxE studies in the order of 500K–1M individuals. Furthermore, the implementation of life-course approaches, or promising and more complex alternative methods to investigate interactions such as Bayesian networks, is even more computationally demanding.

Present and future efforts should also be put into collecting more accurate and reliable phenotypic measures, not only retrospective but longitudinally. Personally, I think that, in a future, data with contextual stress should be extracted and integrated with other sources of data directly from linkable electronic resources. Some information requested in questionnaires about SLE could be extracted directly from national resources without requiring direct reports from participants. New strategies to extract data on life events directly from national population-based electronic records, smart devices and social networks, covering the full spectrum of positive and negative environmental influences, should allow us to construct more replicable measures across studies, with the advantage of reducing sources of bias (e.g. due to self-reporting measures), and longitudinal scores over lifespan. This may establish future resources for GxE studies, with potential to be implemented under a life-course approach and a *diathesis susceptibility* perspective. A life-course approach using linkable resources could improve the detection of GxE effects, while reducing potential confounding effects; improving our understanding of the effects of stress and the etiological mechanisms underpinning MDD, mental health and stress-related disease in general. However, to define and validate the best measures derived from

linkable data in order to cover all contextual stress with effects on MDD is going to be tricky.

In addition, current advances in sequencing technologies allow for affordable whole-genome and whole-exome sequencing that allows the design of well-powered association study of rare (including *de novo* mutations) and structural variants. Wong *et al.* is a nice example of how to take advantage of this new technology with still relative low samples sizes[343]. The potential of new sequencing technologies, added to the increasing sample sizes accompanied with better environmental measures provide the basis for future well-powered GxE research. These strategies should allow us to better understand how an individual's exposure to SLE modulates the effect of genetic risk factors on liability to MDD, and how GxE contribute to its aetiology and "missing heritability"[543]. However, to understand the genetic responses to psychological stress underlying the aetiology of MDD is only one of many strategies to provide insights on the biological mechanism underlying MDD. Eventually, the implementation of such data and knowledge into systems biology approaches would likely enhance our understanding of the disorder by integrating multilevel genomic (e.g. epigenomic, transctiptomic, metabolomics, etc.), environmental and phenotypic data, sophisticated statistical and theoretical methods, and high-performance computing; this is going to allow us to fully determine individual risk, allowing stratifying MDD and providing high-quality level of personalized medicine. Nevertheless, new collaborative efforts and the development of new computational approaches to conduct these high computationally demanding tasks are challenges that must be faced.

Bibliography

1. Organization, W.H. Depression and Other Common Mental Disorders: Global Health Estimates. (Geneva, 2017).
2. Spijker, J. *et al.* Functional disability and depression in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Acta Psychiatr Scand* **110**, 208-14 (2004).
3. Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P. & Jullagate, S. Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatr Dis Treat* **10**, 2097-103 (2014).
4. Moussavi, S. *et al.* Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851-8 (2007).
5. Angst, F., Stassen, H.H., Clayton, P.J. & Angst, J. Mortality of patients with mood disorders: follow-up over 34-38 years. *J Affect Disord* **68**, 167-81 (2002).
6. Veisani, Y., Mohamadian, F. & Delpisheh, A. Prevalence and co-morbidity in common mental disorders and association with suicidal ideation in adult population. *Epidemiol Health* (2017).
7. Kessler, R.C. The costs of depression. *Psychiatr Clin North Am* **35**, 1-14 (2012).
8. Schofield, D.J. *et al.* The indirect economic impacts of co-morbidities on people with depression. *J Psychiatr Res* **47**, 796-801 (2013).
9. Kessler, R.C. & Bromet, E.J. The epidemiology of depression across cultures. *Annu Rev Public Health* **34**, 119-38 (2013).
10. Otte, C. *et al.* Major depressive disorder. *Nature Reviews Disease Primers* **2**, 16065 (2016).
11. Post, R.M. Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. *Am J Psychiatry* **149**, 999-1010 (1992).
12. Plomin, R., Owen, M. & McGuffin, P. The genetic basis of complex human behaviors. *Science* **264**, 1733-1739 (1994).
13. Flint, J. & Kendler, K.S. The genetics of major depression. *Neuron* **81**, 484-503 (2014).
14. McGuffin, P. & Rivera, M. The interaction between stress and genetic factors in the etiopathogenesis of depression. *World Psychiatry* **14**, 161-3 (2015).
15. Kendler, K.S. Explanatory models for psychiatric illness. *Am J Psychiatry* **165**, 695-702 (2008).
16. Association, A.P. Diagnostic and statistical manual of mental disorders (4th ed., Text Revision). Washington, DC: Author. (2000).
17. American Psychiatric, A., American Psychiatric, A. & Force, D.S.M.T. *Diagnostic and statistical manual of mental disorders : DSM-5*, (American Psychiatric Association, Arlington, VA, 2013).
18. Organization, W.H. ICD-10 International Statistical Classification of Diseases and Health Related Problems. (Geneva: World Health Organization, 2004).

19. Duval, F., Lebowitz, B.D. & Macher, J.P. Treatments in depression. *Dialogues Clin Neurosci* **8**, 191-206 (2006).
20. Zimmerman, M., Ellison, W., Young, D., Chelminski, I. & Dalrymple, K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Compr Psychiatry* **56**, 29-34 (2015).
21. Goldberg, D. The heterogeneity of "major depression". *World Psychiatry* **10**, 226-8 (2011).
22. Benazzi, F. Various forms of depression. *Dialogues Clin Neurosci* **8**, 151-61 (2006).
23. Smoller, J.W. Disorders and borders: psychiatric genetics and nosology. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 559-78 (2013).
24. Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I. & Brown, W.A. A systematic review of comparative efficacy of treatments and controls for depression. *PLoS One* **7**, e41778 (2012).
25. Davey, C.G. & Chanen, A.M. The unfulfilled promise of the antidepressant medications. *Med J Aust* **204**, 348-50 (2016).
26. Fournier, J.C. *et al.* Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* **303**, 47-53 (2010).
27. Kirsch, I. *et al.* Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* **5**, e45 (2008).
28. Gibbons, R.D., Hur, K., Brown, C.H., Davis, J.M. & Mann, J.J. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* **69**, 572-9 (2012).
29. Valenstein, M., Vijan, S., Zeber, J.E., Boehm, K. & Buttar, A. The cost-utility of screening for depression in primary care. *Ann Intern Med* **134**, 345-60 (2001).
30. Disease, G.B.D., Injury, I. & Prevalence, C. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545-1602 (2016).
31. Friedrich, M.J. Depression is the leading cause of disability around the world. *JAMA* **317**, 1517-1517 (2017).
32. Ferrari, A.J. *et al.* Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med* **43**, 471-81 (2013).
33. Ferrari, A.J. *et al.* The epidemiological modelling of major depressive disorder: application for the Global Burden of Disease Study 2010. *PLoS One* **8**, e69637 (2013).
34. Luby, J.L. Preschool Depression: The Importance of Identification of Depression Early in Development. *Curr Dir Psychol Sci* **19**, 91-95 (2010).
35. Kessler, R.C. *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **62**, 593-602 (2005).

36. Burcusa, S.L. & Iacono, W.G. Risk for recurrence in depression. *Clin Psychol Rev* **27**, 959-85 (2007).
37. Zisook, S. *et al.* Factors that differentiate early vs. later onset of major depression disorder. *Psychiatry Res* **129**, 127-40 (2004).
38. Liu, Y.H. *et al.* Is early-onset in major depression a predictor of specific clinical features with more impaired social function? *Chin Med J (Engl)* **128**, 811-5 (2015).
39. Zisook, S. *et al.* Effect of age at onset on the course of major depressive disorder. *Am J Psychiatry* **164**, 1539-46 (2007).
40. Weissman, M.M. *et al.* Onset of major depression in early adulthood. Increased familial loading and specificity. *Arch Gen Psychiatry* **41**, 1136-43 (1984).
41. Tozzi, F. *et al.* Family history of depression is associated with younger age of onset in patients with recurrent depression. *Psychol Med* **38**, 641-9 (2008).
42. Lyons, M.J. *et al.* A registry-based twin study of depression in men. *Arch Gen Psychiatry* **55**, 468-72 (1998).
43. Zhu, T. *et al.* Admixture analysis of age at onset in major depressive disorder. *Gen Hosp Psychiatry* **34**, 686-91 (2012).
44. Jaffee, S.R. *et al.* Differences in early childhood risk factors for juvenile-onset and adult-onset depression. *Arch Gen Psychiatry* **59**, 215-22 (2002).
45. Power, R.A. *et al.* Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biol Psychiatry* **81**, 325-335 (2017).
46. Bukh, J.D., Bock, C., Vinberg, M., Gether, U. & Kessing, L.V. Differences between early and late onset adult depression. *Clin Pract Epidemiol Ment Health* **7**, 140-7 (2011).
47. Afifi, M. Gender differences in mental health. *Singapore Med J* **48**, 385-91 (2007).
48. Albert, P.R. Why is depression more prevalent in women? *J Psychiatry Neurosci* **40**, 219-21 (2015).
49. Weissman, M.M. *et al.* Sex differences in rates of depression: cross-national perspectives. *J Affect Disord* **29**, 77-84 (1993).
50. Van de Velde, S., Bracke, P. & Levecque, K. Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Soc Sci Med* **71**, 305-13 (2010).
51. Labonte, B. *et al.* Sex-specific transcriptional signatures in human depression. *Nat Med* (2017).
52. Angst, J. *et al.* Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci* **252**, 201-9 (2002).
53. Piccinelli, M. & Wilkinson, G. Gender differences in depression. Critical review. *Br J Psychiatry* **177**, 486-92 (2000).
54. Kessler, R.C. Epidemiology of women and depression. *J Affect Disord* **74**, 5-13 (2003).

55. Bennett, H.A., Einarson, A., Taddio, A., Koren, G. & Einarson, T.R. Prevalence of depression during pregnancy: systematic review. *Obstet Gynecol* **103**, 698-709 (2004).
56. Lanes, A., Kuk, J.L. & Tamim, H. Prevalence and characteristics of postpartum depression symptomatology among Canadian women: a cross-sectional study. *BMC Public Health* **11**, 302 (2011).
57. O'Hara, M.W. & Swain, A.M. Rates and risk of postpartum depression—a meta-analysis. *International Review of Psychiatry* **8**, 37-54 (1996).
58. Helle, N. *et al.* Very low birth-weight as a risk factor for postpartum depression four to six weeks postbirth in mothers and fathers: Cross-sectional results from a controlled multicentre cohort study. *Journal of Affective Disorders* **180**, 154-161 (2015).
59. Kessler, R.C. *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**, 3095-105 (2003).
60. Kessler, R.C., Merikangas, K.R. & Wang, P.S. Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. *Annu Rev Clin Psychol* **3**, 137-58 (2007).
61. Merikangas, K.R. *et al.* Longitudinal trajectories of depression and anxiety in a prospective community study: the Zurich Cohort Study. *Arch Gen Psychiatry* **60**, 993-1000 (2003).
62. King-Kallimanis, B., Gum, A.M. & Kohn, R. Comorbidity of depressive and anxiety disorders for older Americans in the national comorbidity survey-replication. *Am J Geriatr Psychiatry* **17**, 782-92 (2009).
63. Gao, K. *et al.* Should an assessment of Axis I comorbidity be included in the initial diagnostic assessment of mood disorders? Role of QIDS-16-SR total score in predicting number of Axis I comorbidity. *J Affect Disord* **148**, 256-64 (2013).
64. Donaldson, S.K., Klein, D.N., Riso, L.P. & Schwartz, J.E. Comorbidity between dysthymic and major depressive disorders: a family study analysis. *J Affect Disord* **42**, 103-11 (1997).
65. Kessler, R.C. *et al.* Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *Br J Psychiatry Suppl*, 17-30 (1996).
66. Topic, R. *et al.* Somatic comorbidity, metabolic syndrome, cardiovascular risk, and CRP in patients with recurrent depressive disorders. *Croat Med J* **54**, 453-9 (2013).
67. Lloyd, C.E., Roy, T., Nouwen, A. & Chauhan, A.M. Epidemiology of depression in diabetes: international and cross-cultural issues. *J Affect Disord* **142 Suppl**, S22-9 (2012).
68. Ohayon, M.M. & Schatzberg, A.F. Using chronic pain to predict depressive morbidity in the general population. *Arch Gen Psychiatry* **60**, 39-47 (2003).
69. Slavich, G.M. & Irwin, M.R. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol Bull* **140**, 774-815 (2014).

70. Rosenblat, J.D., Cha, D.S., Mansur, R.B. & McIntyre, R.S. Inflamed moods: a review of the interactions between inflammation and mood disorders. *Prog Neuropsychopharmacol Biol Psychiatry* **53**, 23-34 (2014).
71. Mills, N.T., Scott, J.G., Wray, N.R., Cohen-Woods, S. & Baune, B.T. Research review: the role of cytokines in depression in adolescents: a systematic review. *J Child Psychol Psychiatry* **54**, 816-35 (2013).
72. Pariante, C.M. & Lightman, S.L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci* **31**, 464-8 (2008).
73. Ossowska, K. & Lorenc-Koci, E. Depression in Parkinson's disease. *Pharmacol Rep* **65**, 1545-57 (2013).
74. Kendler, K.S. & Eaves, L.J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am J Psychiatry* **143**, 279-89 (1986).
75. Widom, C.S., DuMont, K. & Czaja, S.J. A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Arch Gen Psychiatry* **64**, 49-56 (2007).
76. Gilman, S.E., Kawachi, I., Fitzmaurice, G.M. & Buka, S.L. Family disruption in childhood and risk of adult depression. *Am J Psychiatry* **160**, 939-46 (2003).
77. Repetti, R.L., Taylor, S.E. & Seeman, T.E. Risky families: family social environments and the mental and physical health of offspring. *Psychol Bull* **128**, 330-66 (2002).
78. Kendler, K.S. & Gardner, C.O. Monozygotic twins discordant for major depression: a preliminary exploration of the role of environmental experiences in the aetiology and course of illness. *Psychol Med* **31**, 411-23 (2001).
79. Kendler, K.S., Kuhn, J. & Prescott, C.A. The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *Am J Psychiatry* **161**, 631-6 (2004).
80. Pakkala, I. *et al.* Genetic contribution to the relationship between personality and depressive symptoms among older women. *Psychol Med* **40**, 1357-66 (2010).
81. Loret de Mola, C., de Franca, G.V., Quevedo Lde, A. & Horta, B.L. Low birth weight, preterm birth and small for gestational age association with adult depression: systematic review and meta-analysis. *Br J Psychiatry* **205**, 340-7 (2014).
82. Polderman, T.J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* **47**, 702-9 (2015).
83. Peyrot, W.J. *et al.* Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry* **205**, 113-9 (2014).
84. Kessler, R.C. The effects of stressful life events on depression. *Annu Rev Psychol* **48**, 191-214 (1997).
85. Tennant, C. Life events, stress and depression: a review of recent findings. *Aust N Z J Psychiatry* **36**, 173-82 (2002).
86. Mullins, N. *et al.* Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med* **46**, 759-70 (2016).

87. Chapman, D.P. *et al.* Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord* **82**, 217-25 (2004).
88. Brooks-Gunn, J. & Duncan, G.J. The effects of poverty on children. *Future Child* **7**, 55-71 (1997).
89. McLeod, J.D. & Shanahan, M.J. Trajectories of poverty and children's mental health. *J Health Soc Behav* **37**, 207-20 (1996).
90. Hammen, C. Stress and depression. *Annu Rev Clin Psychol* **1**, 293-319 (2005).
91. Michael C. Neale, L.R.C. *Methodology for the Study of Twins and Families*, (Kluwer, Dordrecht, the Netherlands, 1992).
92. Smoller, J.W. The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. *Neuropsychopharmacology* **41**, 297-319 (2016).
93. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. Causal relationship between stressful life events and the onset of major depression. *Am J Psychiatry* **156**, 837-41 (1999).
94. Paykel, E.S. Life events and affective disorders. *Acta Psychiatr Scand Suppl*, 61-6 (2003).
95. Stroud, C.B., Davila, J. & Moyer, A. The relationship between stress and depression in first onsets versus recurrences: a meta-analytic review. *J Abnorm Psychol* **117**, 206-13 (2008).
96. Federenko, I.S. *et al.* The heritability of perceived stress. *Psychol Med* **36**, 375-85 (2006).
97. Kendler, K.S. & Baker, J.H. Genetic influences on measures of the environment: a systematic review. *Psychol Med* **37**, 615-26 (2007).
98. Clarke, T.K. *et al.* Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism. *Wellcome Open Res* **3**, 11 (2018).
99. Brugha, T., Bebbington, P., Tennant, C. & Hurry, J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol Med* **15**, 189-94 (1985).
100. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. The assessment of dependence in the study of stressful life events: validation using a twin design. *Psychol Med* **29**, 1455-60 (1999).
101. Silberg, J. *et al.* The influence of genetic factors and life stress on depression among adolescent girls. *Arch Gen Psychiatry* **56**, 225-32 (1999).
102. Kendler, K.S. & Karkowski-Shuman, L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol Med* **27**, 539-47 (1997).
103. Kendler, K.S. *et al.* Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry* **152**, 833-42 (1995).
104. Kendler, K.S. & Gardner, C.O. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. *Arch Gen Psychiatry* **67**, 1120-7 (2010).

105. Harkness, K.L., Monroe, S.M., Simons, A.D. & Thase, M. The generation of life events in recurrent and non-recurrent depression. *Psychol Med* **29**, 135-44 (1999).
106. Harkness, K.L. & Luther, J. Clinical risk factors for the generation of life events in major depression. *J Abnorm Psychol* **110**, 564-72 (2001).
107. Hammen, C. Generation of stress in the course of unipolar depression. *J Abnorm Psychol* **100**, 555-61 (1991).
108. Kendler, K.S., Neale, M., Kessler, R., Heath, A. & Eaves, L. A twin study of recent life events and difficulties. *Arch Gen Psychiatry* **50**, 789-96 (1993).
109. Plomin, R., Lichtenstein, P., Pedersen, N.L., McClearn, G.E. & Nesselroade, J.R. Genetic influence on life events during the last half of the life span. *Psychol Aging* **5**, 25-30 (1990).
110. Bemmels, H.R., Burt, S.A., Legrand, L.N., Iacono, W.G. & McGue, M. The heritability of life events: an adolescent twin and adoption study. *Twin Res Hum Genet* **11**, 257-65 (2008).
111. Boardman, J.D., Alexander, K.B. & Stallings, M.C. Stressful life events and depression among adolescent twin pairs. *Biodemography Soc Biol* **57**, 53-66 (2011).
112. Power, R.A. *et al.* Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol Med* **43**, 1965-71 (2013).
113. Dunn, E.C. *et al.* Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. *Depress Anxiety* **33**, 265-80 (2016).
114. Luciano, M. *et al.* Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. *Intelligence* **38**, 304-313 (2010).
115. Silberg, J., Rutter, M., Neale, M. & Eaves, L. Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *Br J Psychiatry* **179**, 116-21 (2001).
116. Wichers, M. *et al.* Mechanisms of gene-environment interactions in depression: evidence that genes potentiate multiple sources of adversity. *Psychol Med* **39**, 1077-86 (2009).
117. Kraepelin, E. Manic depressive insanity and paranoia. *Edinburgh: E & S Livingston* (1922).
118. Tsuang, M.T. *The genetics of mood disorders / Ming T. Tsuang, Stephen V. Faraone*, (Johns Hopkins University Press, Baltimore, 1990).
119. Sullivan, P.F., Neale, M.C. & Kendler, K.S. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* **157**, 1552-62 (2000).
120. Merikangas, K.R. *et al.* Independence of familial transmission of mania and depression: results of the NIMH family study of affective spectrum disorders. *Mol Psychiatry* **19**, 214-9 (2014).
121. Kendler, K.S., Gatz, M., Gardner, C.O. & Pedersen, N.L. A Swedish national twin study of lifetime major depression. *Am J Psychiatry* **163**, 109-14 (2006).

122. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255-66 (2008).
123. Kendler, K.S., Gardner, C.O., Neale, M.C. & Prescott, C.A. Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychol Med* **31**, 605-16 (2001).
124. Fernandez-Pujals, A.M. *et al.* Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
125. Dawn Teare, M. & Barrett, J.H. Genetic linkage studies. *The Lancet* **366**, 1036-1044 (2005).
126. Mondimore, F.M. & Potash, J.B. Genetic Approaches to Depression: Linkage Studies. in *Biology of Depression: From Novel Insights to Therapeutic Strategies* 735-756 (2008).
127. Tanna, V.L., Winokur, G., Elston, R.C. & Go, R.C. A linkage study of pure depressive disease: the use of the sib-pair method. *Biol Psychiatry* **11**, 767-71 (1976).
128. Abkevich, V. *et al.* Predisposition locus for major depression at chromosome 12q22-12q23.2. *Am J Hum Genet* **73**, 1271-81 (2003).
129. Levinson, D.F. *et al.* Genetics of recurrent early-onset depression (GenRED): design and preliminary clinical characteristics of a repository sample for genetic linkage studies. *Am J Med Genet B Neuropsychiatr Genet* **119B**, 118-30 (2003).
130. Zubenko, G.S. *et al.* Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *Am J Med Genet B Neuropsychiatr Genet* **123B**, 1-18 (2003).
131. Camp, N.J. *et al.* Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *Am J Med Genet B Neuropsychiatr Genet* **135B**, 85-93 (2005).
132. Holmans, P. *et al.* Genetics of recurrent early-onset major depression (GenRED): final genome scan report. *Am J Psychiatry* **164**, 248-58 (2007).
133. Breen, G. *et al.* A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. *Am J Psychiatry* **168**, 840-7 (2011).
134. Pergadia, M.L. *et al.* A 3p26-3p25 genetic linkage finding for DSM-IV major depression in heavy smoking families. *Am J Psychiatry* **168**, 848-52 (2011).
135. Wray, N.R. & Maier, R. Genetic Basis of Complex Genetic Disease: The Contribution of Disease Heterogeneity to Missing Heritability. *Current Epidemiology Reports* **1**, 220-227 (2014).
136. Yu, C. *et al.* Low-frequency and rare variants may contribute to elucidate the genetics of major depressive disorder. *Translational Psychiatry* **8**, 70 (2018).
137. Gatt, J.M., Burton, K.L., Williams, L.M. & Schofield, P.R. Specific and common genes implicated across major mental disorders: a review of meta-analysis studies. *J Psychiatr Res* **60**, 1-13 (2015).

138. Sullivan, P.F. Spurious genetic associations. *Biol Psychiatry* **61**, 1121-6 (2007).
139. Studies, N.-N.W.G.o.R.i.A. *et al.* Replicating genotype-phenotype associations. *Nature* **447**, 655-60 (2007).
140. Bosker, F.J. *et al.* Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry* **16**, 516-32 (2011).
141. Cardon, L.R. Genetics. Delivering new disease genes. *Science* **314**, 1403-5 (2006).
142. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381-5 (2008).
143. Sullivan, P.F. *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* **14**, 359-75 (2009).
144. Sullivan, P.F. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* **68**, 182-6 (2010).
145. Schizophrenia Psychiatric Genome-Wide Association Study, C. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* **43**, 969-76 (2011).
146. Dunn, E.C. *et al.* Genetic determinants of depression: recent findings and future directions. *Harv Rev Psychiatry* **23**, 1-18 (2015).
147. Major Depressive Disorder Working Group of the Psychiatric, G.C. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511 (2013).
148. Hek, K. *et al.* A genome-wide association study of depressive symptoms. *Biol Psychiatry* **73**, 667-78 (2013).
149. consortium, C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588-91 (2015).
150. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
151. Howard, D.M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352 (2019).
152. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
153. Hyde, C.L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**, 1031-6 (2016).
154. Hall, L.S. *et al.* Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Transl Psychiatry* **8**, 9 (2018).
155. Smith, D.J. & Lusis, A.J. The allelic structure of common disease. *Hum Mol Genet* **11**, 2455-61 (2002).
156. Iyengar, S.K. & Elston, R.C. The Genetic Basis of Complex Traits. in *Linkage Disequilibrium and Association Mapping: Analysis and Applications* (ed. Collins, A.R.) 71-84 (Humana Press, Totowa, NJ, 2007).

157. Levinson, D.F. *et al.* Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? *Biol Psychiatry* **76**, 510-2 (2014).
158. Wray, N.R., Lee, S.H. & Kendler, K.S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur J Hum Genet* **20**, 668-74 (2012).
159. Manchia, M. *et al.* The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One* **8**, e76295 (2013).
160. Network & Pathway Analysis Subgroup of Psychiatric Genomics, C. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199-209 (2015).
161. Kao, C.F., Jia, P., Zhao, Z. & Kuo, P.H. Enriched pathways for major depressive disorder identified from a genome-wide association study. *Int J Neuropsychopharmacol* **15**, 1401-11 (2012).
162. Song, G.G., Kim, J.H. & Lee, Y.H. Genome-wide pathway analysis in major depressive disorder. *J Mol Neurosci* **51**, 428-36 (2013).
163. Uher, R. & Zwicker, A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry* **16**, 121-129 (2017).
164. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
165. Wang, T. *et al.* Polygenic risk for five psychiatric disorders and cross-disorder and disorder-specific neural connectivity in two independent populations. *Neuroimage Clin* **14**, 441-449 (2017).
166. Cross-Disorder Group of the Psychiatric Genomics, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).
167. Choi, S.W., Mak, T.S.H. & O'Reilly, P. A guide to performing Polygenic Risk Score analyses. *bioRxiv*, 416545 (2018).
168. Demirkan, A. *et al.* Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry* **16**, 773-83 (2011).
169. Colodro-Conde, L. *et al.* A direct test of the diathesis-stress model for depression. *Mol Psychiatry* (2017).
170. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**, e1003348 (2013).
171. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
172. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
173. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).

174. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
175. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2 (2012).
176. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
177. Lee, S.H. & van der Werf, J.H. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* **32**, 1420-2 (2016).
178. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
179. Hayes, B.J., Visscher, P.M. & Goddard, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res (Camb)* **91**, 47-60 (2009).
180. Meuwissen, T.H., Hayes, B.J. & Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-29 (2001).
181. Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics, C., Wray, N.R. & Lee, S.H. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *Am J Hum Genet* **102**, 1185-1194 (2018).
182. Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984-94 (2013).
183. Ge, T., Chen, C.Y., Neale, B.M., Sabuncu, M.R. & Smoller, J.W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet* **13**, e1006711 (2017).
184. Howard, D.M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun* **9**, 1470 (2018).
185. Gottesman, II & Gould, T.D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**, 636-45 (2003).
186. Glahn, D.C. *et al.* Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. *Am J Med Genet B Neuropsychiatr Genet* **165B**, 122-30 (2014).
187. Iacono, W.G., Vaidyanathan, U., Vrieze, S.I. & Malone, S.M. Knowns and unknowns for psychophysiological endophenotypes: integration and response to commentaries. *Psychophysiology* **51**, 1339-47 (2014).
188. Hasler, G., Drevets, W.C., Manji, H.K. & Charney, D.S. Discovering endophenotypes for major depression. *Neuropsychopharmacology* **29**, 1765-81 (2004).

189. Glahn, D.C. *et al.* High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry* **71**, 6-14 (2012).
190. Beck, A.T., Steer, R.A., & Brown, G.K. Manual for the Beck Depression Inventory-II. *TX: Psychological Corporation* (1996).
191. Eysenck, H.J. & Eysenck, S.B.G. *Manual of the Eysenck personality questionnaire*, (Hodder and Stoughton, London, 1975).
192. Goldstein, B.L. & Klein, D.N. A review of selected candidate endophenotypes for depression. *Clin Psychol Rev* **34**, 417-27 (2014).
193. Rijsdijk, F.V. *et al.* Genetic and environmental influences on psychological distress in the population: General Health Questionnaire analyses in UK twins. *Psychol Med* **33**, 793-801 (2003).
194. Mosing, M.A., Zietsch, B.P., Shekar, S.N., Wright, M.J. & Martin, N.G. Genetic and environmental influences on optimism and its relationship to mental and self-rated health: a study of aging twins. *Behav Genet* **39**, 597-604 (2009).
195. Ayuso-Mateos, J.L., Nuevo, R., Verdes, E., Naidoo, N. & Chatterji, S. From depressive symptoms to depressive disorders: the relevance of thresholds. *Br J Psychiatry* **196**, 365-71 (2010).
196. Kennedy, S.H. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues Clin Neurosci* **10**, 271-7 (2008).
197. Goldberg, D.P. *et al.* The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* **27**, 191-7 (1997).
198. Kendler, K.S. & Myers, J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol Med* **40**, 801-6 (2010).
199. Clark, L.A., Watson, D. & Mineka, S. Temperament, personality, and the mood and anxiety disorders. *J Abnorm Psychol* **103**, 103-16 (1994).
200. Lahey, B.B. Public health significance of neuroticism. *Am Psychol* **64**, 241-56 (2009).
201. Eysenck, S.B.G., Eysenck, H.J. & Barrett, P. A revised version of the psychoticism scale. *Personality and Individual Differences* **6**, 21-29 (1985).
202. Klein, D.N., Kotov, R. & Bufferd, S.J. Personality and depression: explanatory models and review of the evidence. *Annu Rev Clin Psychol* **7**, 269-95 (2011).
203. Wray, N.R., Birley, A.J., Sullivan, P.F., Visscher, P.M. & Martin, N.G. Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res Hum Genet* **10**, 695-702 (2007).
204. van den Berg, S.M. *et al.* Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet* **44**, 295-313 (2014).
205. Lake, R.I., Eaves, L.J., Maes, H.H., Heath, A.C. & Martin, N.G. Further evidence against the environmental transmission of individual differences in neuroticism from

a collaborative study of 45,850 twins and relatives on two continents. *Behav Genet* **30**, 223-33 (2000).
206. Birley, A.J. *et al.* Heritability and nineteen-year stability of long and short EPQ-R Neuroticism scales. *Personality and Individual Differences* **40**, 737-747 (2006).
207. Viken, R.J., Rose, R.J., Kaprio, J. & Koskenvuo, M. A developmental genetic analysis of adult personality: extraversion and neuroticism from 18 to 59 years of age. *J Pers Soc Psychol* **66**, 722-30 (1994).
208. Rettew, D.C. *et al.* The genetic architecture of neuroticism in 3301 Dutch adolescent twins as a function of age and sex: a study from the Dutch twin register. *Twin Res Hum Genet* **9**, 24-9 (2006).
209. Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C. & Eaves, L.J. A longitudinal twin study of personality and major depression in women. *Arch Gen Psychiatry* **50**, 853-62 (1993).
210. Farmer, A. *et al.* Neuroticism, extraversion, life events and depression. The Cardiff Depression Study. *Br J Psychiatry* **181**, 118-22 (2002).
211. Jylha, P. & Isometsa, E. The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depress Anxiety* **23**, 281-9 (2006).
212. Jardine, R., Martin, N.G. & Henderson, A.S. Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet Epidemiol* **1**, 89-107 (1984).
213. Hirschfeld, R.M. & Klerman, G.L. Personality attributes and affective disorders. *Am J Psychiatry* **136**, 67-70 (1979).
214. Hirschfeld, R.M. *et al.* Assessing personality: effects of the depressive state on trait measurement. *Am J Psychiatry* **140**, 695-9 (1983).
215. De Fruyt, F., Van Leeuwen, K., Bagby, R.M., Rolland, J.P. & Rouillon, F. Assessing and interpreting personality change and continuity in patients treated for major depression. *Psychol Assess* **18**, 71-80 (2006).
216. G Bazana, P. & M Stelmack, R. *Stability of Personality Across the Life Span: A Meta-Analysis*, 113-144 (2004).
217. McGue, M., Bacon, S. & Lykken, D.T. Personality Stability and Change in Early Adulthood. *Developmental Psychology* **29**, 96-109 (1993).
218. Jeronimus, B.F., Ormel, J., Aleman, A., Penninx, B.W. & Riese, H. Negative and positive life events are associated with small but lasting change in neuroticism. *Psychol Med* **43**, 2403-15 (2013).
219. Jeronimus, B.F., Riese, H., Sanderman, R. & Ormel, J. Mutual reinforcement between neuroticism and life experiences: a five-wave, 16-year study to test reciprocal causation. *J Pers Soc Psychol* **107**, 751-64 (2014).
220. Riese, H. *et al.* Timing of Stressful Life Events Affects Stability and Change of Neuroticism. *European Journal of Personality* **28**, 193-200 (2014).

221. Tak, L.M., Kingma, E.M., van Ockenburg, S.L., Ormel, J. & Rosmalen, J.G. Age- and sex-specific associations between adverse life events and functional bodily symptoms in the general population. *J Psychosom Res* **79**, 112-6 (2015).
222. Hovens, J.G., Giltay, E.J., van Hemert, A.M. & Penninx, B.W. Childhood Maltreatment and the Course of Depressive and Anxiety Disorders: The Contribution of Personality Characteristics. *Depress Anxiety* **33**, 27-34 (2016).
223. Gautam, S. & Kamal, P. A study of impact of stressful life-events in neurotic patients. *Indian J Psychiatry* **32**, 356-61 (1990).
224. Magnus, K., Diener, E., Fujita, F. & Pavot, W. Extraversion and neuroticism as predictors of objective life events: a longitudinal analysis. *J Pers Soc Psychol* **65**, 1046-53 (1993).
225. Monroe, S.M. & Simons, A.D. Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. *Psychol Bull* **110**, 406-25 (1991).
226. Belsky, J. & Pluess, M. Beyond diathesis stress: differential susceptibility to environmental influences. *Psychol Bull* **135**, 885-908 (2009).
227. Belsky, J., Bakermans-Kranenburg, M.J. & van Ijzendoorn, M.H. For Better and for Worse: Differential Susceptibility to Environmental Influences. *Current Directions in Psychological Science* **16**, 300-304 (2007).
228. Meehl, P.E. Schizotaxia, schizotypy, schizophrenia. *American Psychologist* **17**, 827-838 (1962).
229. Belsky, J. & Beaver, K.M. Cumulative-genetic plasticity, parenting and adolescent self-regulation. *J Child Psychol Psychiatry* **52**, 619-26 (2011).
230. Belsky, J. *et al.* Vulnerability genes or plasticity genes? *Mol Psychiatry* **14**, 746-54 (2009).
231. Pluess, M. & Belsky, J. Vantage sensitivity: individual differences in response to positive experiences. *Psychol Bull* **139**, 901-16 (2013).
232. de Villiers, B., Lionetti, F. & Pluess, M. Vantage sensitivity: a framework for individual differences in response to psychological intervention. *Social Psychiatry and Psychiatric Epidemiology* **53**, 545-554 (2018).
233. Kraft, P., Yen, Y.C., Stram, D.O., Morrison, J. & Gauderman, W.J. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered* **63**, 111-9 (2007).
234. Caspi, A. & Moffitt, T.E. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat Rev Neurosci* **7**, 583-90 (2006).
235. Moffitt, T.E., Caspi, A. & Rutter, M. Measured Gene-Environment Interactions in Psychopathology: Concepts, Research Strategies, and Implications for Research, Intervention, and Public Understanding of Genetics. *Perspect Psychol Sci* **1**, 5-27 (2006).
236. Dick, D.M. Gene-environment interaction in psychological traits and disorders. *Annu Rev Clin Psychol* **7**, 383-409 (2011).

237. Kendler, K.S. Twin studies of psychiatric illness: Current status and future directions. *Archives of General Psychiatry* **50**, 905-915 (1993).
238. Rowe, D.C., Jacobson, K.C. & Van den Oord, E.J. Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Dev* **70**, 1151-62 (1999).
239. Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B. & Gottesman, II. Socioeconomic status modifies heritability of IQ in young children. *Psychol Sci* **14**, 623-8 (2003).
240. Ellis, B.J. & Boyce, W.T. Biological Sensitivity to Context. *Current Directions in Psychological Science* **17**, 183-187 (2008).
241. Essex, M.J., Armstrong, J.M., Burk, L.R., Goldsmith, H.H. & Boyce, W.T. Biological sensitivity to context moderates the effects of the early teacher-child relationship on the development of mental health by adolescence. *Dev Psychopathol* **23**, 149-61 (2011).
242. Ellis, B.J., Essex, M.J. & Boyce, W.T. Biological sensitivity to context: II. Empirical explorations of an evolutionary-developmental theory. *Dev Psychopathol* **17**, 303-28 (2005).
243. Jaffee, S.R. & Price, T.S. Genotype-environment correlations: implications for determining the relationship between environmental exposures and psychiatric illness. *Psychiatry* **7**, 496-499 (2008).
244. Lindstrom, S., Yen, Y.C., Spiegelman, D. & Kraft, P. The impact of gene-environment dependence and misclassification in genetic association studies incorporating gene-environment interactions. *Hum Hered* **68**, 171-81 (2009).
245. Lesch, K.P. Gene-environment interaction and the genetics of depression. *J Psychiatry Neurosci* **29**, 174-84 (2004).
246. Thapar, A., Harold, G. & McGuffin, P. Life Events and Depressive Symptoms in Childhood — Shared Genes or Shared Adversity? A Research Note. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* **39**, 1153-1158 (1998).
247. McGuffin, P., Katz, R. & Bebbington, P. The Camberwell Collaborative Depression Study. III. Depression and adversity in the relatives of depressed probands. *Br J Psychiatry* **152**, 775-82 (1988).
248. Farmer, A. *et al.* Cardiff depression study. A sib-pair study of life events and familiality in major depression. *Br J Psychiatry* **176**, 150-5 (2000).
249. Caspi, A. *et al.* Role of genotype in the cycle of violence in maltreated children. *Science* **297**, 851-4 (2002).
250. Caspi, A. *et al.* Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301**, 386-9 (2003).
251. Munafo, M.R., Durrant, C., Lewis, G. & Flint, J. Gene X environment interactions at the serotonin transporter locus. *Biol Psychiatry* **65**, 211-9 (2009).
252. Risch, N. *et al.* Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* **301**, 2462-71 (2009).

253. Karg, K., Burmeister, M., Shedden, K. & Sen, S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry* **68**, 444-54 (2011).
254. Duncan, L.E., Pollastri, A.R. & Smoller, J.W. Mind the gap: why many geneticists and psychological scientists have discrepant views about gene-environment interaction (GxE) research. *Am Psychol* **69**, 249-68 (2014).
255. Halldorsdottir, T. & Binder, E.B. Gene x Environment Interactions: From Molecular Mechanisms to Behavior. *Annu Rev Psychol* **68**, 215-241 (2017).
256. Mandelli, L. & Serretti, A. Gene environment interaction studies in depression and suicidal behavior: An update. *Neurosci Biobehav Rev* **37**, 2375-97 (2013).
257. Dunn, E.C. *et al.* Research review: gene-environment interaction research in youth depression - a systematic review with recommendations for future research. *J Child Psychol Psychiatry* **52**, 1223-38 (2011).
258. Mandelli, L., Petrelli, C. & Serretti, A. The role of specific early trauma in adult depression: A meta-analysis of published literature. Childhood trauma and adult depression. *Eur Psychiatry* **30**, 665-80 (2015).
259. Uher, R. Gene-environment interactions in common mental disorders: an update and strategy for a genome-wide search. *Soc Psychiatry Psychiatr Epidemiol* **49**, 3-14 (2014).
260. Cohen-Woods, S., Craig, I.W. & McGuffin, P. The current state of play on the molecular genetics of depression. *Psychol Med* **43**, 673-87 (2013).
261. Caspi, A., Hariri, A.R., Holmes, A., Uher, R. & Moffitt, T.E. Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. *Am J Psychiatry* **167**, 509-27 (2010).
262. Culverhouse, R.C. *et al.* Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol Psychiatry* **23**, 133-142 (2018).
263. Thomas, D. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet* **11**, 259-72 (2010).
264. Karg, K. & Sen, S. Gene x environment interaction models in psychiatric genetics. *Curr Top Behav Neurosci* **12**, 441-62 (2012).
265. Borglum, A.D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol Psychiatry* **19**, 325-33 (2014).
266. Avramopoulos, D. *et al.* Infection and inflammation in schizophrenia and bipolar disorder: a genome wide study for interactions with genetic variation. *PLoS One* **10**, e0116696 (2015).
267. Otowa, T. *et al.* The First Pilot Genome-Wide Gene-Environment Study of Depression in the Japanese Population. *PLoS One* **11**, e0160823 (2016).
268. Ikeda, M. *et al.* Genome-wide environment interaction between depressive state and stressful life events. *J Clin Psychiatry* **77**, e29-30 (2016).

269. Coleman, J.R.I., Eley, T.C. & Breen, G. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *bioRxiv* (2018).
270. Van der Auwera, S. *et al.* Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes: The childhood trauma working-group of PGC-MDD. *Am J Med Genet B Neuropsychiatr Genet* **177**, 40-49 (2018).
271. Khoury, M.J. & Wacholder, S. Invited Commentary: From Genome-Wide Association Studies to Gene-Environment-Wide Interaction Studies—Challenges and Opportunities. *American Journal of Epidemiology* **169**, 227-230 (2009).
272. Almli, L.M. *et al.* Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392-9 (2014).
273. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
274. Aschard, H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol* **40**, 678-688 (2016).
275. Peyrot, W.J. *et al.* Does Childhood Trauma Moderate Polygenic Risk for Depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biol Psychiatry* (2017).
276. Musliner, K.L. *et al.* Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychol Med* **45**, 1709-20 (2015).
277. Monroe, S.M. Modern approaches to conceptualizing and measuring human life stress. *Annu Rev Clin Psychol* **4**, 33-52 (2008).
278. Porcelli, B. *et al.* Association between stressful life events and autoimmune diseases: A systematic review and meta-analysis of retrospective case-control studies. *Autoimmun Rev* **15**, 325-34 (2016).
279. Lin, Y. *et al.* Striking life events associated with primary breast cancer susceptibility in women: a meta-analysis study. *J Exp Clin Cancer Res* **32**, 53 (2013).
280. Jafri, S.H.R. *et al.* Major stressful life events and risk of developing lung cancer. *Journal of Clinical Oncology* **35**, 1575-1575 (2017).
281. Bogdan, R., Nikolova, Y.S. & Pizzagalli, D.A. Neurogenetics of depression: a focus on reward processing and stress sensitivity. *Neurobiol Dis* **52**, 12-23 (2013).
282. Lazarus, R.S. & Folkman, S. Transactional theory and research on emotions and coping. *European Journal of Personality* **1**, 141-169 (1987).
283. Smith, M.A. *et al.* The relationship between Type D personality and physical health complaints is mediated by perceived stress and anxiety but not diurnal cortisol secretion. *Stress*, 1-8 (2018).
284. Herr, R.M. *et al.* Long-Term Effectiveness of Stress Management at Work: Effects of the Changes in Perceived Stress Reactivity on Mental Health and Sleep Problems Seven Years Later. *Int J Environ Res Public Health* **15**(2018).

285. Moore, R.C. *et al.* Complex interplay between health and successful aging: role of perceived stress, resilience, and social support. *Am J Geriatr Psychiatry* **23**, 622-32 (2015).
286. Hayman, L.W., Jr., Lucas, T. & Porcerelli, J.H. Cognitive appraisal vs. exposure-based stress measures: links to perceived mental and physical health in low-income black women. *J Nerv Ment Dis* **202**, 807-12 (2014).
287. Harris, M.L., Loxton, D., Sibbritt, D.W. & Byles, J.E. The influence of perceived stress on the onset of arthritis in women: findings from the Australian Longitudinal Study on women's health. *Ann Behav Med* **46**, 9-18 (2013).
288. Rietschel, L. *et al.* Perceived stress has genetic influences distinct from neuroticism and depression. *Behav Genet* **44**, 639-45 (2014).
289. Rietschel, L. *et al.* Hair Cortisol in Twins: Heritability and Genetic Overlap with Psychological Variables and Stress-System Genes. *Sci Rep* **7**, 15351 (2017).
290. Bleys, D., Luyten, P., Soenens, B. & Claes, S. Gene-environment interactions between stress and 5-HTTLPR in depression: A meta-analytic update. *Journal of Affective Disorders* **226**, 339-345 (2018).
291. Vukasovic, T. & Bratko, D. Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol Bull* **141**, 769-85 (2015).
292. Levinson, D.F. The genetics of depression: a review. *Biol Psychiatry* **60**, 84-92 (2006).
293. Middeldorp, C.M., Cath, D.C., Van Dyck, R. & Boomsma, D.I. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med* **35**, 611-24 (2005).
294. Smith, D.J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* **21**, 749-57 (2016).
295. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* **50**, 6-11 (2018).
296. Smits, D.J.M. & Boeck, P.D. From BIS/BAS to the big five. *European Journal of Personality* **20**, 255-270 (2006).
297. Schneider, T.R., Rench, T.A., Lyons, J.B. & Riffle, R.R. The influence of neuroticism, extraversion and openness on stress responses. *Stress Health* **28**, 102-10 (2012).
298. Kim, S.E. *et al.* Direct and Indirect Effects of Five Factor Personality and Gender on Depressive Symptoms Mediated by Perceived Stress. *PLoS One* **11**, e0154140 (2016).
299. Bazana, P.G. & Stelmack, R.M. Chapter 8 - Stability of Personality Across the Life Span: A Meta-Analysis. in *On the Psychobiology of Personality* 113-144 (Elsevier, Oxford, 2004).
300. Nivard, M.G., Middeldorp, C.M., Dolan, C.V. & Boomsma, D.I. Genetic and Environmental Stability of Neuroticism From Adolescence to Adulthood. *Twin Res Hum Genet* **18**, 746-54 (2015).

301. Biobank, U. UK Biobank: protocol for a large-scale prospective epidemiological resource. (2010).
302. Biobank, U. UK Biobank ethics and governance framework, version 3.0 (2007).
303. Biobank, U. Genotype imputation and genetic association studies of UK Biobank. (2015).
304. Amador, C. *et al.* Recent genomic heritage in Scotland. *BMC Genomics* **16**, 437 (2015).
305. Smith, B.H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* **42**, 689-700 (2013).
306. Smith, B.H. *et al.* Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* **7**, 74 (2006).
307. Kerr, S.M. *et al.* Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med Genet* **14**, 38 (2013).
308. Biobank, U. Touchscreen questionnaire. (2012).
309. Eysenck HJ, E.S. Manual of the Eysenck Personality Questionnaire. (Hodder and Stoughton, London, 1975).
310. Smith, D.J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362 (2013).
311. Spitzer, R.L., Kroenke, K. & Williams, J.B. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* **282**, 1737-44 (1999).
312. First MB, S.R., Gibbon M, Williams JB. Structured Clinical Interview for DSM-IV-TR Axis I Disorders. (New York State Psychiatric Institute, New York, 2002).
313. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-8 (2015).
314. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
315. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
316. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
317. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
318. Mailman, M.D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* **39**, 1181-6 (2007).
319. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
320. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).

321. Genetics of Personality, C. *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **72**, 642-50 (2015).
322. Team, R.C. *R: A language and environment for statistical computing*, (R Foundation for Statistical Computing, Vienna, Austria, 2015).
323. Chen, G.B. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Front Genet* **5**, 107 (2014).
324. Hirschfeld, R.M. *et al.* Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry* **157**, 1873-5 (2000).
325. Raine, A. The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophr Bull* **17**, 555-64 (1991).
326. Zuo, L. *et al.* A New Genomewide Association Meta-Analysis of Alcohol Dependence. *Alcohol Clin Exp Res* **39**, 1388-95 (2015).
327. Cook, M.N. *et al.* Identification of candidate genes that underlie the QTL on chromosome 1 that mediates genetic differences in stress-ethanol interactions. *Physiol Genomics* **47**, 308-17 (2015).
328. Zuo, L. *et al.* Common PTP4A1-PHF3-EYS variants are specific for alcohol dependence. *Am J Addict* **23**, 411-4 (2014).
329. Zuo, L., Zhang, X., Deng, H.W. & Luo, X. Association of rare PTP4A1-PHF3-EYS variants with alcohol dependence. *J Hum Genet* **58**, 178-9 (2013).
330. Zuo, L. *et al.* A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. *PLoS One* **6**, e26726 (2011).
331. Hontelez, S., Karthaus, N., Looman, M.W., Ansems, M. & Adema, G.J. DC-SCRIPT regulates glucocorticoid receptor function and expression of its target GILZ in dendritic cells. *J Immunol* **190**, 3172-9 (2013).
332. Lopez-Garcia, J. *et al.* ZNF366 is an estrogen receptor corepressor that acts through CtBP and histone deacetylases. *Nucleic Acids Res* **34**, 6126-36 (2006).
333. Miyamoto-Sato, E. *et al.* A comprehensive resource of interacting protein regions for refining human transcription factor networks. *PLoS One* **5**, e9289 (2010).
334. Aragam, N., Wang, K.S. & Pan, Y. Genome-wide association analysis of gender differences in major depressive disorder in the Netherlands NESDA and NTR population-based samples. *J Affect Disord* **133**, 516-21 (2011).
335. Fidalgo, T.M., da Silveira, E.D. & da Silveira, D.X. Psychiatric comorbidity related to alcohol use among adolescents. *Am J Drug Alcohol Abuse* **34**, 83-9 (2008).
336. Lopez, B., Turner, R.J. & Saavedra, L.M. Anxiety and risk for substance dependence among late adolescents/young adults. *J Anxiety Disord* **19**, 275-94 (2005).
337. Low, N.C., Lee, S.S., Johnson, J.G., Williams, J.B. & Harris, E.S. The association between anxiety and alcohol versus cannabis abuse disorders among adolescents in primary care settings. *Fam Pract* **25**, 321-7 (2008).

338. Rutledge, P.C. & Sher, K.J. Heavy drinking from the freshman year into early young adulthood: the roles of stress, tension-reduction drinking motives, gender and personality. *J Stud Alcohol* **62**, 457-66 (2001).
339. Schmidt, N.B., Buckner, J.D. & Keough, M.E. Anxiety sensitivity as a prospective predictor of alcohol use disorders. *Behav Modif* **31**, 202-19 (2007).
340. Varlinskaya, E.I., Kim, E.U. & Spear, L.P. Chronic intermittent ethanol exposure during adolescence: Effects on stress-induced social alterations and social drinking in adulthood. *Brain Res* **1654**, 145-156 (2017).
341. Torigata, K. *et al.* LATS2 Positively Regulates Polycomb Repressive Complex 2. *PLoS One* **11**, e0158562 (2016).
342. Shen, E.Y. *et al.* Neuronal Deletion of Kmt2a/Mll1 Histone Methyltransferase in Ventral Striatum is Associated with Defective Spike-Timing-Dependent Striatal Synaptic Plasticity, Altered Response to Dopaminergic Drugs, and Increased Anxiety. *Neuropsychopharmacology* **41**, 3103-3113 (2016).
343. Wong, M.L. *et al.* The PHF21B gene is associated with major depression and modulates the stress response. *Mol Psychiatry* **22**, 1015-1025 (2017).
344. Walsh, R.M. *et al.* Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. *Nat Commun* **8**, 15142 (2017).
345. Gage, S.H. *et al.* Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomization. *Sci Rep* **7**, 40653 (2017).
346. Hartz, S.M. *et al.* Genetic correlation between smoking behaviors and schizophrenia. *Schizophr Res* (2017).
347. Smith, S.M. & Vale, W.W. The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. *Dialogues Clin Neurosci* **8**, 383-95 (2006).
348. Holsboer, F. The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology* **23**, 477-501 (2000).
349. Mello, A.F., Mello, M.F., Carpenter, L.L. & Price, L.H. Update on stress and depression: the role of the hypothalamic-pituitary-adrenal (HPA) axis. *Braz J Psychiatr* **25**, 231-8 (2003).
350. Sapolsky, R.M., Romero, L.M. & Munck, A.U. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr Rev* **21**, 55-89 (2000).
351. Herbert, J. *et al.* Do corticosteroids damage the brain? *J Neuroendocrinol* **18**, 393-411 (2006).
352. Bellavance, M.A. & Rivest, S. The HPA - Immune Axis and the Immunomodulatory Actions of Glucocorticoids in the Brain. *Front Immunol* **5**, 136 (2014).
353. Turnbull, A.V. & Rivier, C.L. Regulation of the hypothalamic-pituitary-adrenal axis by cytokines: actions and mechanisms of action. *Physiol Rev* **79**, 1-71 (1999).
354. Raison, C.L., Capuron, L. & Miller, A.H. Cytokines sing the blues: inflammation and the pathogenesis of depression. *Trends Immunol* **27**, 24-31 (2006).

355. Lasic, D. *et al.* Metabolic syndrome and inflammation markers in patients with schizophrenia and recurrent depressive disorder. *Psychiatr Danub* **26**, 214-9 (2014).
356. Heim, C., Newport, D.J., Bonsall, R., Miller, A.H. & Nemeroff, C.B. Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *Am J Psychiatry* **158**, 575-81 (2001).
357. Lopez-Duran, N.L., Kovacs, M. & George, C.J. Hypothalamic-pituitary-adrenal axis dysregulation in depressed children and adolescents: a meta-analysis. *Psychoneuroendocrinology* **34**, 1272-83 (2009).
358. Lopez-Duran, N.L. *et al.* HPA-axis stress reactivity in youth depression: evidence of impaired regulatory processes in depressed boys. *Stress* **18**, 545-53 (2015).
359. Guerry, J.D. & Hastings, P.D. In search of HPA axis dysregulation in child and adolescent depression. *Clin Child Fam Psychol Rev* **14**, 135-60 (2011).
360. Mello, A.F. *et al.* [Depression and stress: is there an endophenotype?]. *Braz J Psychiatr* **29 Suppl 1**, S13-8 (2007).
361. Shea, A., Walsh, C., Macmillan, H. & Steiner, M. Child maltreatment and HPA axis dysregulation: relationship to major depressive disorder and post traumatic stress disorder in females. *Psychoneuroendocrinology* **30**, 162-78 (2005).
362. Stephens, M.A. & Wand, G. Stress and the HPA axis: role of glucocorticoids in alcohol dependence. *Alcohol Res* **34**, 468-83 (2012).
363. Schepis, T.S., Rao, U., Yadav, H. & Adinoff, B. The limbic-hypothalamic-pituitary-adrenal axis and the development of alcohol use disorders in youth. *Alcohol Clin Exp Res* **35**, 595-605 (2011).
364. Dostert, A. & Heinzel, T. Negative glucocorticoid receptor response elements and their role in glucocorticoid action. *Curr Pharm Des* **10**, 2807-16 (2004).
365. Del Monaco, M. *et al.* Identification of Novel Glucocorticoid-Response Elements in Human Elastin Promoter and Demonstration of Nucleotide Sequence Specificity of the Receptor Binding. *Journal of Investigative Dermatology* **108**, 938-942 (1997).
366. Glucocorticoid Response Elements (GRE). in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* 803-803 (Springer Netherlands, Dordrecht, 2008).
367. Reddy, T.E. *et al.* Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* **19**, 2163-71 (2009).
368. Polman, J.A. *et al.* A genome-wide signature of glucocorticoid receptor binding in neuronal PC12 cells. *BMC Neurosci* **13**, 118 (2012).
369. Yu, C.Y. *et al.* Genome-wide analysis of glucocorticoid receptor binding regions in adipocytes reveal gene network involved in triglyceride homeostasis. *PLoS One* **5**, e15188 (2010).
370. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**, 264-8 (2011).
371. Sheikh, H.I., Kryski, K.R., Smith, H.J., Hayden, E.P. & Singh, S.M. Corticotropin-releasing hormone system polymorphisms are associated with children's cortisol reactivity. *Neuroscience* **229**, 1-11 (2013).

372. Mahon, P.B., Zandi, P.P., Potash, J.B., Nestadt, G. & Wand, G.S. Genetic association of FKBP5 and CRHR1 with cortisol response to acute psychosocial stress in healthy adults. *Psychopharmacology (Berl)* **227**, 231-41 (2013).
373. Zannas, A.S. & Binder, E.B. Gene-environment interactions at the FKBP5 locus: sensitive periods, mechanisms and pleiotropism. *Genes Brain Behav* **13**, 25-37 (2014).
374. DeRijk, R.H. *et al.* A common polymorphism in the mineralocorticoid receptor modulates stress responsiveness. *J Clin Endocrinol Metab* **91**, 5083-9 (2006).
375. Feurer, C. *et al.* HPA axis multilocus genetic profile score moderates the impact of interpersonal stress on prospective increases in depressive symptoms for offspring of depressed mothers. *J Abnorm Psychol* **126**, 1017-1028 (2017).
376. Di Iorio, C.R. *et al.* Hypothalamic-pituitary-adrenal axis genetic variation and early stress moderates amygdala function. *Psychoneuroendocrinology* **80**, 170-178 (2017).
377. Arloth, J. *et al.* Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. *Neuron* **86**, 1189-202 (2015).
378. Arnau-Soler, A. *et al.* Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS One* **13**, e0209160 (2018).
379. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res* **43**, D662-9 (2015).
380. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **43**, W589-98 (2015).
381. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
382. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
383. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288-9 (2009).
384. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
385. Maglott, D., Ostell, J., Pruitt, K.D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **35**, D26-31 (2007).
386. Koch, C.E., Leinweber, B., Drengberg, B.C., Blaum, C. & Oster, H. Interaction between circadian rhythms and stress. *Neurobiol Stress* **6**, 57-67 (2017).
387. Lyall, L.M. *et al.* Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *Lancet Psychiatry* **5**, 507-514 (2018).
388. Turecki, G. & Meaney, M.J. Effects of the Social Environment and Stress on Glucocorticoid Receptor Gene Methylation: A Systematic Review. *Biol Psychiatry* **79**, 87-96 (2016).

389. Efstathopoulos, P. *et al.* NR3C1 hypermethylation in depressed and bullied adolescents. *Translational Psychiatry* **8**, 121 (2018).
390. Pagliaccio, D. *et al.* Stress-System Genes and Life Stress Predict Cortisol Levels and Amygdala and Hippocampal Volumes in Children. *Neuropsychopharmacology* **39**, 1245 (2013).
391. Gatt, J.M. *et al.* Interactions between BDNF Val66Met polymorphism and early life stress predict brain and arousal pathways to syndromal depression and anxiety. *Mol Psychiatry* **14**, 681-95 (2009).
392. Aguilera, M. *et al.* Early adversity and 5-HTT/BDNF genes: new evidence of gene–environment interactions on depressive symptoms in a general population. *Psychological Medicine* **39**, 1425-1432 (2009).
393. Hosang, G.M., Shiles, C., Tansey, K.E., McGuffin, P. & Uher, R. Interaction between stress and the BDNF Val66Met polymorphism in depression: a systematic review and meta-analysis. *BMC Med* **12**, 7 (2014).
394. Newell-Price, J., Bertagna, X., Grossman, A.B. & Nieman, L.K. Cushing's syndrome. *The Lancet* **367**, 1605-1617 (2006).
395. Thomsen, A.F., Kvist, T.K., Andersen, P.K. & Kessing, L.V. The risk of affective disorders in patients with adrenocortical insufficiency. *Psychoneuroendocrinology* **31**, 614-22 (2006).
396. Velders, F.P. *et al.* Genetics of cortisol secretion and depressive symptoms: a candidate gene and genome wide association approach. *Psychoneuroendocrinology* **36**, 1053-61 (2011).
397. van der Meer, D. *et al.* Predicting attention-deficit/hyperactivity disorder severity from psychosocial stress and stress-response genes: a random forest regression approach. *Transl Psychiatry* **7**, e1145 (2017).
398. Cicchetti, D., Rogosch, F.A. & Sturge-Apple, M.L. Interactions of child maltreatment and serotonin transporter and monoamine oxidase A polymorphisms: depressive symptomatology among adolescents from low socioeconomic status backgrounds. *Dev Psychopathol* **19**, 1161-80 (2007).
399. Melas, P.A. *et al.* Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities. *Int J Neuropsychopharmacol* **16**, 1513-28 (2013).
400. Ensel, W.M., Peek, M.K., Lin, N. & Lai, G. Stress in the life course: a life history approach. *J Aging Health* **8**, 389-416 (1996).
401. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. *J Nerv Ment Dis* **186**, 661-9 (1998).
402. Mazure, C.M. Life Stressors as Risk Factors in Depression. *Clinical Psychology: Science and Practice* **5**, 291-313 (1998).
403. Lichtenberg, P. & Belmaker, R.H. Subtyping major depressive disorder. *Psychother Psychosom* **79**, 131-5 (2010).

404. Elisei, S., Sciarma, T., Verdolini, N. & Anastasi, S. Resilience and depressive disorders. *Psychiatr Danub* **25 Suppl 2**, S263-7 (2013).
405. Vogel, F. Schizophrenia genesis: The origins of madness. *American Journal of Human Genetics* **48**, 1218-1218 (1991).
406. Mann, J.J., Waternaux, C., Haas, G.L. & Malone, K.M. Toward a clinical model of suicidal behavior in psychiatric patients. *Am J Psychiatry* **156**, 181-9 (1999).
407. Riemann, D. *et al.* The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med Rev* **14**, 19-31 (2010).
408. Bolt, M.A., Helming, L.M. & Tintle, N.L. The Associations between Self-Reported Exposure to the Chernobyl Nuclear Disaster Zone and Mental Health Disorders in Ukraine. *Front Psychiatry* **9**, 32 (2018).
409. Iyegbe, C., Campbell, D., Butler, A., Ajnakina, O. & Sham, P. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc Psychiatry Psychiatr Epidemiol* **49**, 169-82 (2014).
410. McGrath, J.J., Mortensen, P.B., Visscher, P.M. & Wray, N.R. Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. *Schizophr Bull* **39**, 955-9 (2013).
411. Plomin, R. Commentary: missing heritability, polygenic scores, and gene-environment correlation. *J Child Psychol Psychiatry* **54**, 1147-9 (2013).
412. Wray, N.R. *et al.* Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* **55**, 1068-87 (2014).
413. Gunderson, K.L. Whole-Genome Genotyping on Bead Arrays. in *DNA Microarrays for Biomedical Research: Methods and Protocols* (ed. Dufva, M.) 197-213 (Humana Press, Totowa, NJ, 2009).
414. Nagy, R. *et al.* Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med* **9**, 23 (2017).
415. Navrady, L.B. *et al.* Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *Int J Epidemiol* (2017).
416. Goldberg, D.P. & Hillier, V.F. A scaled version of the General Health Questionnaire. *Psychol Med* **9**, 139-45 (1979).
417. Sterling, M. General Health Questionnaire - 28 (GHQ-28). *J Physiother* **57**, 259 (2011).
418. Banks, M.H. Validation of the General Health Questionnaire in a young community sample. *Psychol Med* **13**, 349-53 (1983).
419. Marks, A.D.G., Horrocks, K.A. & Schutte, N.S. Emotional intelligence mediates the relationship between insecure attachment and subjective health outcomes. *Personality and Individual Differences* **98**, 188-192 (2016).

420. O'Rourke, S., MacHale, S., Signorini, D. & Dennis, M. Detecting psychiatric morbidity after stroke: comparison of the GHQ and the HAD Scale. *Stroke* **29**, 980-5 (1998).
421. Kessler, R.C., Andrews, G., Mroczek, D., Ustun, B. & Wittchen, H.-U. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *International Journal of Methods in Psychiatric Research* **7**, 171-185 (1998).
422. Brugha, T.S. & Cragg, D. The List of Threatening Experiences: the reliability and validity of a brief life events questionnaire. *Acta Psychiatr Scand* **82**, 77-81 (1990).
423. Motrico, E. *et al.* Psychometric properties of the List of Threatening Experiences--LTE and its association with psychosocial factors and mental disorders according to different scoring methods. *J Affect Disord* **150**, 931-40 (2013).
424. Keller, M.C. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* **75**, 18-24 (2014).
425. Eaves, L.J., Last, K., Martin, N.G. & Jinks, J.L. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology* **30**, 1-42 (1977).
426. Plomin, R., DeFries, J.C. & Loehlin, J.C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* **84**, 309-22 (1977).
427. Arnau-Soler, A. *et al.* Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Transl Psychiatry* **9**, 14 (2019).
428. Vrshek-Schallhorn, S. *et al.* Additive genetic risk from five serotonin system polymorphisms interacts with interpersonal stress to predict depression. *J Abnorm Psychol* **124**, 776-90 (2015).
429. Li, J.J., Berk, M.S. & Lee, S.S. Differential susceptibility in longitudinal models of gene-environment interaction for adolescent depression. *Dev Psychopathol* **25**, 991-1003 (2013).
430. Kang, S.-M. & G. Waller, N. *Moderated Multiple Regression, Spurious Interaction Effects, and IRT*, 87-105 (2005).
431. Arnau-Soler, A. *et al.* A validation of the diathesis-stress model for depression in Generation Scotland. *Transl Psychiatry* **9**, 25 (2019).
432. Garantziotis, S. & Schwartz, D.A. Ecogenomics of respiratory diseases of public health significance. *Annu Rev Public Health* **31**, 37-51 1 p following 51 (2010).
433. Aschard, H. *et al.* Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol* **46**, 894-904 (2017).
434. Molfino, N.A. & Coyle, A.J. Gene-environment interactions in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* **3**, 491-7 (2008).
435. Polonikov, A.V., Ivanov, V.P. & Solodilova, M.A. Genetic variation of genes for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the importance of

gene-gene and gene-environment interactions for disease susceptibility. *J Hum Genet* **54**, 440-9 (2009).
436. Haiman, C.A. *et al.* Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med* **354**, 333-42 (2006).
437. Han, J., Hankinson, S.E., Colditz, G.A. & Hunter, D.J. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *Br J Cancer* **91**, 1604-9 (2004).
438. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
439. Wang, L., Murk, W. & DeWan, A.T. Genome-Wide Gene by Environment Interaction Analysis Identifies Common SNPs at 17q21.2 that Are Associated with Increased Body Mass Index Only among Asthmatics. *PLoS One* **10**, e0144114 (2015).
440. Siegert, S. *et al.* Genome-wide investigation of gene-environment interactions in colorectal cancer. *Hum Genet* **132**, 219-31 (2013).
441. Gong, J. *et al.* Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. *PLoS Genet* **12**, e1006296 (2016).
442. Polfus, L.M. *et al.* Genome-wide association study of gene by smoking interactions in coronary artery calcification. *PLoS One* **8**, e74642 (2013).
443. Duncan, L.E. & Keller, M.C. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry* **168**, 1041-9 (2011).
444. Salleh, M.R. Life event, stress and illness. *Malays J Med Sci* **15**, 9-18 (2008).
445. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
446. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
447. Wang, L. *et al.* [Value of patient health questionnaires (PHQ)-9 and PHQ-2 for screening depression disorders in cardiovascular outpatients]. *Zhonghua Xin Xue Guan Bing Za Zhi* **43**, 428-31 (2015).
448. Liu, J.Z., Erlich, Y. & Pickrell, J.K. Case-control association mapping by proxy using family history of disease. *Nat Genet* **49**, 325-331 (2017).
449. Altman, D.G. & Royston, P. The cost of dichotomising continuous variables. *BMJ* **332**, 1080 (2006).
450. Stein, J.L. *et al.* Voxelwise genome-wide association study (vGWAS). *Neuroimage* **53**, 1160-74 (2010).
451. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet* **49**, 1576-1583 (2017).
452. Hart, A.B. *et al.* Genome-wide association study of d-amphetamine response in healthy volunteers identifies putative associations, including cadherin 13 (CDH13). *PLoS One* **7**, e42646 (2012).

453. Sugimoto, K. *et al.* The induction of H3K9 methylation by PIWIL4 at the p16Ink4a locus. *Biochem Biophys Res Commun* **359**, 497-502 (2007).
454. Sivagurunathan, S., Arunachalam, J.P. & Chidambaram, S. PIWI-like protein, HIWI2 is aberrantly expressed in retinoblastoma cells and affects cell-cycle potentially through OTX2. *Cell Mol Biol Lett* **22**, 17 (2017).
455. Lee, H.H.C. *et al.* Genetic Otx2 mis-localization delays critical period plasticity across brain regions. *Mol Psychiatry* **22**, 680-688 (2017).
456. Sonuga-Barke, E.J. *et al.* Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 1359-68 (2008).
457. Belzeaux, R. *et al.* Responder and nonresponder patients exhibit different peripheral transcriptional signatures during major depressive episode. *Transl Psychiatry* **2**, e185 (2012).
458. Mamdani, F., Berlim, M.T., Beaulieu, M.M. & Turecki, G. Pharmacogenomic predictors of citalopram treatment outcome in major depressive disorder. *World J Biol Psychiatry* **15**, 135-44 (2014).
459. Shi, G. *et al.* PHD finger protein 2 (PHF2) represses ribosomal RNA gene transcription by antagonizing PHF finger protein 8 (PHF8) and recruiting methyltransferase SUV39H1. *J Biol Chem* **289**, 29691-700 (2014).
460. Yu, D. *et al.* Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *Am J Psychiatry* **172**, 82-93 (2015).
461. Ceccato, S., Kudielka, B.M. & Schwieren, C. Increased Risk Taking in Relation to Chronic Stress in Adults. *Front Psychol* **6**, 2036 (2015).
462. Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A. & Spinath, F.M. Life events as environmental States and genetic traits and the role of personality: a longitudinal twin study. *Behav Genet* **42**, 57-72 (2012).
463. Conway, C.C., Hammen, C., Brennan, P.A., Lind, P.A. & Najman, J.M. Interaction of chronic stress with serotonin transporter and catechol-O-methyltransferase polymorphisms in predicting youth depression. *Depress Anxiety* **27**, 737-45 (2010).
464. Cicchetti, D. & Rogosch, F.A. Genetic moderation of child maltreatment effects on depression and internalizing symptoms by serotonin transporter linked polymorphic region (5-HTTLPR), brain-derived neurotrophic factor (BDNF), norepinephrine transporter (NET), and corticotropin releasing hormone receptor 1 (CRHR1) genes in African American children. *Dev Psychopathol* **26**, 1219-39 (2014).
465. Rutter, M., Moffitt, T.E. & Caspi, A. Gene-environment interplay and psychopathology: multiple varieties but real effects. *J Child Psychol Psychiatry* **47**, 226-61 (2006).
466. Maier, R.M., Visscher, P.M., Robinson, M.R. & Wray, N.R. Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychol Med* **48**, 1055-1067 (2018).

467. Ahmad, S. *et al.* Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. *PLoS Genet* **9**, e1003607 (2013).
468. Qi, Q. *et al.* Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med* **367**, 1387-96 (2012).
469. Uher, R. Gene-environment interactions in severe mental illness. *Front Psychiatry* **5**, 48 (2014).
470. Keers, R. & Uher, R. Gene-environment interaction in major depression and antidepressant treatment response. *Curr Psychiatry Rep* **14**, 129-37 (2012).
471. Gonda, X. *et al.* Significance of risk polymorphisms for depression depends on stress exposure. *Scientific Reports* **8**, 3946 (2018).
472. Kishi, T., Yoshimura, R., Ikuta, T. & Iwata, N. Brain-Derived Neurotrophic Factor and Major Depressive Disorder: Evidence from Meta-Analyses. *Front Psychiatry* **8**, 308 (2017).
473. van Winkel, M. *et al.* Impact of variation in the BDNF gene on social stress sensitivity and the buffering impact of positive emotions: replication and extension of a gene-environment interaction. *Eur Neuropsychopharmacol* **24**, 930-8 (2014).
474. Ghinelli, E. *et al.* Presence and localization of neurotrophins and neurotrophin receptors in rat lacrimal gland. *Invest Ophthalmol Vis Sci* **44**, 3352-7 (2003).
475. Kafitz, K.W., Rose, C.R., Thoenen, H. & Konnerth, A. Neurotrophin-evoked rapid excitation through TrkB receptors. *Nature* **401**, 918-21 (1999).
476. Jang, K.L., Livesley, W.J. & Vernon, P.A. Heritability of the big five personality dimensions and their facets: a twin study. *J Pers* **64**, 577-91 (1996).
477. Huls, A. *et al.* Comparison of weighting approaches for genetic risk scores in gene-environment interaction studies. *BMC Genet* **18**, 115 (2017).
478. Scarr, S. & McCartney, K. How people make their own environments: a theory of genotype greater than environment effects. *Child Dev* **54**, 424-35 (1983).
479. Starr, L.R., Hammen, C., Conway, C.C., Raposa, E. & Brennan, P.A. Sensitizing effect of early adversity on depressive reactions to later proximal stress: Moderation by polymorphisms in serotonin transporter and corticotropin releasing hormone receptor genes in a 20-year longitudinal study. *Dev Psychopathol* **26**, 1241-54 (2014).
480. Keers, R. & Pluess, M. Childhood quality influences genetic sensitivity to environmental influences across adulthood: A life-course Gene x Environment interaction study. *Dev Psychopathol* **29**, 1921-1933 (2017).
481. Assary, E., Vincent, J.P., Keers, R. & Pluess, M. Gene-environment interaction and psychiatric disorders: Review and future directions. *Semin Cell Dev Biol* (2017).
482. Aiken, L.S. & West, S.G. *Multiple regression: Testing and interpreting interactions*, xi, 212-xi, 212 (Sage Publications, Inc, Thousand Oaks, CA, US, 1991).
483. Greenland, S. Tests for interaction in epidemiologic studies: a review and a study of power. *Stat Med* **2**, 243-51 (1983).
484. McClelland, G.H. & Judd, C.M. Statistical difficulties of detecting interactions and moderator effects. *Psychol Bull* **114**, 376-90 (1993).

485. Wong, M.Y., Day, N.E., Luan, J.A., Chan, K.P. & Wareham, N.J. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int J Epidemiol* **32**, 51-7 (2003).
486. Eley, T.C. *et al.* Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Mol Psychiatry* **9**, 908-15 (2004).
487. Uddin, M. *et al.* Gender differences in the genetic and environmental determinants of adolescent depression. *Depress Anxiety* **27**, 658-66 (2010).
488. Gonda, X. *et al.* Genetic variants in major depressive disorder: From pathophysiology to therapy. *Pharmacol Ther* (2018).
489. Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* **167**, 748-51 (2010).
490. Lupien, S.J. *et al.* The DSM5/RDoC debate on the future of mental health research: implication for studies on human stress and presentation of the signature bank. *Stress* **20**, 2-18 (2017).
491. Triantis, V. *et al.* Identification and characterization of DC-SCRIPT, a novel dendritic cell-expressed member of the zinc finger family of transcriptional regulators. *J Immunol* **176**, 1081-9 (2006).
492. T., K. Glucocorticoid Receptor. .
493. Ancelin, M.L. & Ryan, J. 5-HTTLPR × stress hypothesis: is the debate over? *Molecular Psychiatry* (2017).
494. Clarke, T.K. *et al.* HPA-axis activity in alcoholism: examples for a gene-environment interaction. *Addict Biol* **13**, 1-14 (2008).
495. Liu, J. *et al.* Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. *Neuropsychopharmacology* **31**, 1574-82 (2006).
496. Schol-Gelok, S. *et al.* A genome-wide screen for depression in two independent Dutch populations. *Biol Psychiatry* **68**, 187-96 (2010).
497. O'Donovan, M.C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* **40**, 1053-5 (2008).
498. McGue, M. *et al.* A genome-wide association study of behavioral disinhibition. *Behav Genet* **43**, 363-73 (2013).
499. Gianoulakis, C. Influence of the endogenous opioid system on high alcohol consumption and genetic predisposition to alcoholism. *J Psychiatry Neurosci* **26**, 304-18 (2001).
500. Foster, D.W. *et al.* Tears in your beer: Gender differences in coping drinking motives, depressive symptoms and drinking. *Int J Ment Health Addict* **12**, 730-746 (2014).
501. Kim, J.E., Song, I.H. & Lee, S.H. Gender Differences of Stressful Life Events, Coping Style, Symptom Severity, and Health-Related Quality of Life in Patients With Panic Disorder. *J Nerv Ment Dis* **205**, 714-719 (2017).
502. Mariotti, A. The effects of chronic stress on health: new insights into the molecular mechanisms of brain-body communication. *Future Sci OA* **1**, FSO23 (2015).

503. Hakimi, M.A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* **99**, 7420-5 (2002).
504. Klajn, A. *et al.* The rest repression of the neurosecretory phenotype is negatively modulated by BHC80, a protein of the BRAF/HDAC complex. *J Neurosci* **29**, 6296-307 (2009).
505. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718-22 (2007).
506. Kang, M.K., Mehrazarin, S., Park, N.H. & Wang, C.Y. Epigenetic gene regulation by histone demethylases: emerging role in oncogenesis and inflammation. *Oral Dis* **23**, 709-720 (2017).
507. Park, S.Y., Park, J.W. & Chun, Y.S. Jumonji histone demethylases as emerging therapeutic targets. *Pharmacol Res* **105**, 146-51 (2016).
508. Sanchez, R. & Zhou, M.M. The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364-72 (2011).
509. Musselman, C.A. & Kutateladze, T.G. PHD fingers: epigenetic effectors and potential drug targets. *Mol Interv* **9**, 314-23 (2009).
510. Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell Death Discov* **3**, 17071 (2017).
511. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22 (2017).
512. Vicente, C.T., Revez, J.A. & Ferreira, M.A.R. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology* **6**, e165 (2017).
513. Torkamani, A., Wineinger, N.E. & Topol, E.J. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* **19**, 581-590 (2018).
514. Lewis, C.M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Med* **9**, 96 (2017).
515. Zheutlin, A.B. & Ross, D.A. Polygenic Risk Scores: What Are They Good For? *Biol Psychiatry* **83**, e51-e53 (2018).
516. Palk, A.C., Dalvie, S., de Vries, J., Martin, A.R. & Stein, D.J. Potential use of clinical polygenic risk scores in psychiatry - ethical implications and communicating high polygenic risk. *Philos Ethics Humanit Med* **14**, 4 (2019).
517. Wray, N.R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14**, 507-15 (2013).
518. Hewitt, J.K. Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav Genet* **42**, 1-2 (2012).
519. Johnson, E.O. *et al.* Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence. *Addiction* **105**, 2014-22 (2010).
520. Thomas, D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annu Rev Public Health* **31**, 21-36 (2010).

521. Luan, J.A., Wong, M.Y., Day, N.E. & Wareham, N.J. Sample size determination for studies of gene-environment interaction. *Int J Epidemiol* **30**, 1035-40 (2001).
522. Xu, M.K. *et al.* Psychometric precision in phenotype definition is a useful step in molecular genetic investigation of psychiatric disorders. *Transl Psychiatry* **5**, e593 (2015).
523. Rosenman, R., Tennekoon, V. & Hill, L.G. Measuring bias in self-reported data. *Int J Behav Healthc Res* **2**, 320-332 (2011).
524. Dempfle, A. *et al.* Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur J Hum Genet* **16**, 1164-72 (2008).
525. Smith, G.D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1-22 (2003).
526. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* **9**, 224 (2018).
527. Hammen, C., Henry, R. & Daley, S.E. Depression and sensitization to stressors among young women as a function of childhood adversity. *J Consult Clin Psychol* **68**, 782-7 (2000).
528. Kendler, K.S., Kuhn, J.W. & Prescott, C.A. Childhood sexual abuse, stressful life events and risk for major depression in women. *Psychol Med* **34**, 1475-82 (2004).
529. Espejo, E.P. *et al.* Stress sensitization and adolescent depressive severity as a function of childhood adversity: a link to anxiety disorders. *J Abnorm Child Psychol* **35**, 287-99 (2007).
530. Hazel, N.A., Hammen, C., Brennan, P.A. & Najman, J. Early childhood adversity and adolescent depression: the mediating role of continued stress. *Psychol Med* **38**, 581-9 (2008).
531. McLaughlin, K.A., Conron, K.J., Koenen, K.C. & Gilman, S.E. Childhood adversity, adult stressful life events, and risk of past-year psychiatric disorder: a test of the stress sensitization hypothesis in a population-based sample of adults. *Psychol Med* **40**, 1647-58 (2010).
532. St Clair, M.C. *et al.* Childhood adversity subtypes and depressive symptoms in early and late adolescence. *Dev Psychopathol* **27**, 885-99 (2015).
533. Hankin, B.L. & Abramson, L.Y. Development of gender differences in depression: an elaborated cognitive vulnerability-transactional stress theory. *Psychol Bull* **127**, 773-96 (2001).
534. Rudolph, K.D. & Flynn, M. Childhood adversity and youth depression: influence of gender and pubertal status. *Dev Psychopathol* **19**, 497-521 (2007).
535. Pagliaccio, D. & Barch, D.M. Chapter 2 - Early Life Adversity and Risk for Depression: Alterations in Cortisol and Brain Structure and Function as Mediating Mechanisms. in *Systems Neuroscience in Depression* (ed. Frodl, T.) 29-77 (Academic Press, San Diego, 2016).

536. McGonagle, K.A. & Kessler, R.C. Chronic stress, acute stress, and depressive symptoms. *Am J Community Psychol* **18**, 681-706 (1990).
537. Ruch, L.O., Chandler, S.M. & Harter, R.A. Life change and rape impact. *J Health Soc Behav* **21**, 248-60 (1980).
538. Harkness, K.L., Bruce, A.E. & Lumley, M.N. The role of childhood abuse and neglect in the sensitization to stressful life events in adolescent depression. *J Abnorm Psychol* **115**, 730-41 (2006).
539. Kendler, K.S., Thornton, L.M. & Gardner, C.O. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the "kindling" hypothesis. *Am J Psychiatry* **157**, 1243-51 (2000).
540. Brown, G.W. & Harris, T.O. Depression and the serotonin transporter 5-HTTLPR polymorphism: A review and a hypothesis concerning gene–environment interaction. *Journal of Affective Disorders* **111**, 1-12 (2008).
541. Gibson, G. Decanalization and the origin of complex disease. *Nat Rev Genet* **10**, 134-40 (2009).
542. Gillum, L.A. *et al.* NIH disease funding levels and burden of disease. *PLoS One* **6**, e16837 (2011).
543. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**, 446-50 (2010).

Appendices

Appendix A

Appendix A contains supplementary material for **chapter 2**: Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. The published article in *PLOS ONE* is also included.

A.1 DEPICT analyses

Gene sets were analysed using DEPICT (https://github.com/perslab/depict)[1] to (i) prioritise genes in independent loci, (ii) identify reconstituted gene sets enriched by genes selected, which may represent biologically relevant pathways and systems, and (iii) determine enriched tissue/cell types.

SNPs from meta-analyzed GWIS with stress-sensitivity (SS) effect with $p < 2x10^{-5}$ (see **Appendix A.5** Supplementary Figure 2) were clumped using PLINK v1.9[2] to identify 12 independently associated "lead SNPs" (LD r2 > 0.1; physical kb threshold = 500kb; 1000 Genomes Project Phase 1 CEU, GBR, TSI genotype data[3]). Associated regions were defined by linkage disequilibrium (LD) around the 12 "lead SNPs" (LD $r^2 > 0.5$; 1000 Genomes Project Phase 1 CEU, GBR, TSI genotype data) and genes were selected if they mapped within or overlapping the regions identified (genome build GRCh37). Genes within the high LD HLA locus (chr6:25000000-35000000) were removed and overlapping regions merged. If no gene was present in a region, the nearest gene was selected. 13 unique genes were finally selected. By comparing these associated regions with genome-wide randomly-selected loci and matched for gene density, DEPICT determined whether these genes share biological function, based on the hypothesis that genes truly associated with stress-sensitivity will be part of the same mechanisms underlying this trait. No significant pathway or mechanism was found at FDR < 0.05. DEPICT is based on predicted function of genes derived using the results of 77,840 microarrays from two human, one rat and one mouse Affymetrix gene

expression platforms from the Omnibus (GeO) database[4], each covering expression of 19,997 genes.

A.2 Polygenic risk profiling and MDD models

PRS weighted by SS effect ($\hat{\beta}_{SS}$) for each individual on GS:SFHS were estimated using GWIS statistics from UKB as follows,

$$(i) \quad PRS_{SS} = \sum_{j=1}^{m} \hat{\beta}_{SSj} SNP_j$$

Using MDD-GWAS statistics from UKB (discovery sample), we estimated for each SNP (ii) the main additive effect on MDD and (iii) the main additive effect on EPQN, from the following additive genetic models,

$$(ii) \quad MDD = \beta_{Di} SNP_i + COVARIATES + \varepsilon$$

$$(iii) \quad EPQN = \beta_{Ni} SNP_i + COVARIATES + \varepsilon$$

Where $i \, \epsilon \, \{1 \ldots n\}$; n = total number of SNPs on UKB sample (n = 557,813). Using these effects, we created MDD and EPQN PRS for each individual weighting by β_D (PRS_D) and β_N PRS_N on GS:SFHS (target sample) as follows,

$$(ii) \quad PRS_D = \sum_{k=1}^{l} \hat{\beta}_{Dk} SNP_k$$

$$(iii) \quad PRS_N = \sum_{p=1}^{t} \hat{\beta}_{Np} SNP_p$$

Where $k \, \epsilon \, \{1 \ldots l\}$; l ≤ n; l = number of SNPs at best MDD prediction fit in GS:SFHS and $p \, \epsilon \, \{1 \ldots t\}$; t ≤ n; t = number of SNPs at best EPQN prediction fit in GS:SFHS.

All PRS at best fit (i.e. PRS_{SS}, PRS_{D} and PRS_{N}) were combined on several general linear models to assess MDD status (case-control) prediction on GS:SFHS as follows,

null model: $MDD \sim COVARIATES$

model 1: $MDD \sim PRS_{SS} + COVARIATES$

model 2: $MDD \sim PRS_{D} + COVARIATES$

model 3: $MDD \sim PRS_{N} + COVARIATES$

model 4: $MDD \sim PRS_{D} + PRS_{N} + COVARIATES$

model 5: $MDD \sim PRS_{D} + PRS_{SS} + COVARIATES$

model 6: $MDD \sim PRS_{N} + PRS_{SS} + COVARIATES$

full model: $MDD \sim PRS_{SS} + PRS_{D} + PRS_{N} + COVARIATES$

Before determining the scores, strand-ambiguous SNPs were removed from the genotype data. SNPs present in both the discovery and target samples were clumped to obtain a set of independent SNPs in approximate linkage equilibrium ($r^2 < 0.1$, within a 250kb window). PRS were generated for up to 13 *p* thresholds (< 0.001, < 0.005, < 0.01, < 0.02, < 0.03, < 0.04, < 0.05, < 0.1, < 0.2, < 0.3, < 0.4, < 0.5, <=1). Scores were standardized to a mean of 0 and a standard deviation of 1 for use in further analyses. Each score was regressed on MDD status using logistic regression models adjusted for sex, age and 20 PCs and permuted 10,000 times to assess association with MDD status. Nagelkerke's R^2 coefficients, a likelihood-based measure extensively used in prediction of psychiatric disorders[5,6] reflecting the proportion of MDD risk explained by each model at the observed scale, were calculated and converted into R^2 coefficients at the liability scale using Hong Lee's transformation[7] available from GEAR: GEnetic Analysis Repository[8]. To assess MDD risk explained at the population level, we used prevalence of 12.2% in GS:SFHS[9] and 25.8% in UKB[10]. Significance of each PRS was assesses by likelihood ratio test. Cross-validation was performed following the

same procedure above using GS:SFHS as discovery sample and UKB as target sample to predict MDD phenotype (dependent variable) under a quasi-binomial distribution after being pre-adjusted by centre, array and genotyping batch as random effects, in a general linear regression model adjusting by sex, age and 15 PCs. Finally, the analysis was replicated and cross-validated as detailed above using summary statistics from the most recent Psychiatric Genetic Consortium MDD meta-analysis and the Genetics of Personality Consortium neuroticism meta-analysis to weight PRS_D and PRS_N by the main MDD and neuroticism additive effects, respectively.

A.3 References

1. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
2. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
3. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
4. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**, D991-5 (2013).
5. Cross-Disorder Group of the Psychiatric Genomics, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).
6. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
7. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
8. Chen, G.B. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Front Genet* **5**, 107 (2014).
9. Fernandez-Pujals, A.M. *et al.* Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
10. Smith, D.J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362 (2013).

A.4 Supplementary Tables

Supplementary Table 1 EPQN comparison between MDD cases and healthy controls

	GS:SFHS (N = 7 155)					
	Unmatched samples			Matched samples		
	Controls	MDD cases	*p* value	Controls	MDD cases	*p* value
n	5 145	2 010	-	2 010	2 010	-
Sex = females (%)	2095 (40.7)	1340 (66.7)	< 0.001	1 329 (66.1)	1 340 (66.7)	0.738
Age mean (s.d.)	51.80 (11.78)	46.99 (12.06)	< 0.001	47.52 (11.90)	46.99 (12.06)	0.155
PRS_D mean (s.d.)	-0.04 (0.68)	0.09 (0.69)	< 0.001	0.09 (0.66)	0.09 (0.69)	0.99
EPQN mean (s.d.)	3.16 (2.89)	6.42 (3.32)	< 0.001	3.56 (3.00)	6.42 (3.32)	< 0.001

	UKB (N = 23 092)					
	Unmatched samples			Matched samples		
	Controls	MDD cases	*p* value	Controls	MDD cases	*p* value
n	15 258	7 834	-	7 834	7 834	-
Sex = females (%)	6 654 (43.6)	4 856 (62.0)	< 0.001	4 857 (62.0)	4 856 (62.0)	1
Age mean (s.d.)	58.30 (8.06)	56.64 (7.92)	< 0.001	56.71 (8.23)	56.64 (7.92)	0.631
PRS_D mean (s.d.)	-0.06 (0.99)	0.11 (1.00)	< 0.001	0.11 (0.98)	0.11 (1.00)	0.542
EPQN mean (s.d.)	2.79 (2.62)	5.64 (3.30)	< 0.001	2.97 (2.66)	5.64 (3.30)	< 0.001

Matched samples: cases and controls were matched for PRS weighted by MDD-PGC GWAS (PRS_D), sex and age**.**

Supplementary Table 2 Top 10 SNPs from GWIS on UK Biobank

CHR	SNP	BP	A1	BETA1	SE1	BETA2	SE2	BETA	*p* value	P.TYPE	GENE	G.POSITION
11	**rs2221540**	**132716369**	**G**	**-0.1362**	**0.06346**	**0.5033**	**0.1096**	**0.6395**	**$4.41x10^{-7}$**	**intronic**	**OPCML**	**-**
6	**rs319924**	**64487247**	**G**	**0.08516**	**0.03036**	**-0.2229**	**0.05302**	**-0.30806**	**$4.62x10^{-7}$**	**intronic**	**EYS**	**-**
10	rs2256220	24856314	C	-0.07145	0.03121	0.2416	0.05459	0.31305	$6.43x10^{-7}$	intergenic	ARHGAP21	16224
16	rs4783107	84996447	T	-0.1134	0.04602	0.3404	0.08248	0.4538	$1.55x10^{-6}$	intergenic	ZDHHC7	13502
10	rs2265265	24854876	T	-0.06464	0.03012	0.2206	0.05277	0.28524	$2.68x10^{-6}$	intergenic	ARHGAP21	17662
19	**rs1078734**	**39308874**	**C**	**0.08035**	**0.03003**	**-0.2066**	**0.05328**	**-0.28695**	**$2.71x10^{-6}$**	**intronic**	**ECH1/HNRNPL**	**-**
11	**rs4575282**	**132719646**	**C**	**-0.1084**	**0.06247**	**0.4692**	**0.107**	**0.5776**	**$3.16x10^{-6}$**	**intronic**	**OPCML**	**-**
6	rs1057530	64427095	T	0.0948	0.03001	-0.1876	0.05283	-0.2824	$3.34x10^{-6}$	intergenic	PHF3	-3491
6	**rs10485358**	**64386060**	**T**	**0.09146**	**0.03004**	**-0.1902**	**0.05277**	**-0.28166**	**$3.51x10^{-6}$**	**intronic**	**PHF3**	**-**
5	rs3913723	41560079	G	0.1371	0.06798	-0.4449	0.1153	-0.582	$1.37x10^{-5}$	intergenic	PLCXD3	-49451

A1 minor allele, BETA1 SNP effect in neuroticism score (EPQN) within MDD controls, BETA2 SNP effect in EPQN within MDD cases, BETA stress-sensitivity effect (i.e. differential allelic effect in EPQN between MDD cases and controls; BETA2 effect minus BETA1 effect), p value reflects significance of BETA coefficient, P.TYPE SNP's position type, GENE closest gene within 100kb, G.POSITION position of the closest gene (bp). In bold hits from intragenic SNPs.

Supplementary Table 3 Top 10 SNPs from GWIS on Generation Scotland: Scottish Family Health Study

CHR	SNP	BP	A1	BETA1	SE1	BETA2	SE2	BETA	*p* value	P.TYPE	GENE	G.POSITION
3	rs10510908	63762994	A	-0.07792	0.02704	0.1588	0.04224	0.23672	$2.36x10^{-6}$	intergenic	THOC7	56302
11	rs7934698	134412739	A	-0.0633	0.03157	0.2141	0.04959	0.2774	$2.37x10^{-6}$	intergenic	-	-
2	**rs7606549**	**115490198**	**A**	**0.02374**	**0.02006**	**-0.154**	**0.03187**	**-0.17774**	**$2.37x10^{-6}$**	**intronic**	**DPP10**	**-**
4	rs6851779	45487679	A	0.09384	0.03685	-0.2235	0.05645	-0.31734	$2.51x10^{-6}$	intergenic	-	-
2	**rs319858**	**115530841**	**G**	**0.05248**	**0.01983**	**-0.1204**	**0.03181**	**-0.17288**	**$4.00x10^{-6}$**	**intronic**	**DPP10**	**-**
3	rs13098181	63749405	A	-0.07987	0.02695	0.1452	0.04225	0.22507	$7.11x10^{-6}$	intergenic	THOC7	13358
3	rs12108177	63673795	G	-0.07852	0.0261	0.1383	0.04087	0.21682	$7.76x10^{-6}$	intergenic	SNTN	-9349
20	rs6063840	51208949	G	0.06519	0.02116	-0.1145	0.03416	-0.17969	$7.77x10^{-6}$	intergenic	LINC01524	28178
12	**rs4765723**	**3352543**	**C**	**-0.04887**	**0.02382**	**0.1471**	**0.03699**	**0.19597**	**$8.39x10^{-6}$**	**intronic**	**TSPAN9**	**-**
20	rs6068235	51167169	A	0.06096	0.02135	-0.1195	0.03443	-0.18046	$8.42x10^{-6}$	intergenic	LINC01524	69958

A1 minor allele, P.TYPE SNP's position type, GENE closest gene within 100kb, G.POSITION position of the closest gene (bp). In bold hits from intragenic SNPs.

Supplementary Table 4 Traits with significant evidence of association with closest gene to suggestive stress-sensitive hits

The closest genes to SNPs associated with stress-sensitivity at suggestive significance levels have prior evidence of association in dbGAP with a wide range of neuropsychiatric traits such as schizophrenia, bipolar disorder, attention deficit disorder with hyperactivity, mental competency, intuition, sleep or alcohol drinking.

Trait	SNP	Chr	Location	Context	Gene	*p* value	Source	PubMed
Cholesterol, LDL	rs10515153	5	71989150	intergenic	ZNF366	2.18×10^{-25}	dbGaP	17903299
Alcohol Drinking	rs10515153	5	71989150	intergenic	ZNF366	3.50×10^{-15}	dbGaP	0
Cholesterol, LDL	rs10515153	5	71989150	intergenic	ZNF366	9.38×10^{-12}	dbGaP	17903299
Echocardiography	rs10515153	5	71989150	intergenic	ZNF366	4.78×10^{-08}	dbGaP	17903301
Mental Competency	rs3858216	10	24689550	intron	KIAA1217	7.72×10^{-16}	dbGaP	17903295
Respiratory Function Tests	rs1941212	11	133288335	intron	OPCML	4.91×10^{-10}	dbGaP	17903307
Respiratory Function Tests	rs1941212	11	133288335	intron	OPCML	8.44×10^{-10}	dbGaP	17903307
Respiratory Function Tests	rs1941212	11	133288335	intron	OPCML	4.58×10^{-08}	dbGaP	17903307
Respiratory Function Tests	rs1941211	11	133288419	intron	OPCML	5.78×10^{-12}	dbGaP	17903307
Respiratory Function Tests	rs1941211	11	133288419	intron	OPCML	1.18×10^{-11}	dbGaP	17903307
Exercise Test	rs1941211	11	133288419	intron	OPCML	4.56×10^{-11}	dbGaP	17903301
Body Mass Index	rs1941211	11	133288419	intron	OPCML	3.22×10^{-10}	dbGaP	17903300
Cardiovascular Diseases	rs1941211	11	133288419	intron	OPCML	6.05×10^{-10}	dbGaP	17903304
Cardiovascular Diseases	rs1941211	11	133288419	intron	OPCML	9.36×10^{-10}	dbGaP	17903304
Body Mass Index	rs1941211	11	133288419	intron	OPCML	1.27×10^{-09}	dbGaP	17903300
Angiography	rs1941211	11	133288419	intron	OPCML	2.65×10^{-09}	dbGaP	17903301
Cardiovascular Diseases	rs1941211	11	133288419	intron	OPCML	2.81×10^{-09}	dbGaP	17903304
Angiography	rs1941211	11	133288419	intron	OPCML	1.54×10^{-08}	dbGaP	17903301
Body Mass Index	rs1941211	11	133288419	intron	OPCML	2.03×10^{-08}	dbGaP	17903300
Respiratory Function Tests	rs1941211	11	133288419	intron	OPCML	2.08×10^{-08}	dbGaP	17903307

Respiratory Function Tests	rs1941211	11	133288419	intron	OPCML	3.05 x10^{-08}	dbGaP	17903307
Cholesterol, HDL	rs10507129	12	101478127	intron	ANO4	1.44 x10^{-10}	dbGaP	17903299
Cardiovascular Diseases	rs2278075	16	78291173	intron	WWOX	5.32 x10^{-09}	dbGaP	17903304
Body Mass Index	rs9319533	16	79088978	intron	WWOX	3.56 x10^{-11}	dbGaP	17903300
C-Reactive Protein	rs9319533	16	79088978	intron	WWOX	6.00 x10^{-11}	dbGaP	17903293
Insulin	rs9319533	16	79088978	intron	WWOX	6.89 x10^{-10}	dbGaP	17903298
Body Weight	rs9319533	16	79088978	intron	WWOX	9.54 x10^{-10}	dbGaP	17903300
Body Mass Index	rs9319533	16	79088978	intron	WWOX	2.82 x10^{-09}	dbGaP	17903300

Trait	**SNP**	**CHR**	**BP**	**Context**	**Gene**	***p* value**	**Source**	**PubMed**
Mental Competency	rs3858216	10	24689550	intron	KIAA1217	7.72 x10^{-16}	dbGaP	17903295
Cardiovascular Diseases	rs1941211	11	133288419	intron	OPCML	6.05 x10^{-10}	dbGaP	17903304
Sleep	rs1939966	11	132297204	intron	OPCML	7.82 x10^{-06}	dbGaP	17903308
Diabetes Mellitus	rs7927989	11	133356488	intron	OPCML	7.89 x10^{-06}	dbGaP	0
Schizophrenia	rs7125438	11	132402910	intron	OPCML	2.08 x10^{-05}	dbGaP	0
Attention Deficit Disorder with Hyperactivity	rs10786284	10	98135505	intron	TLL2	2.00 x10^{-06}	NHGRI	18839057
Cardiovascular Diseases	rs2278075	16	78291173	intron	WWOX	5.32 x10^{-09}	dbGaP	17903304
Intuition	rs17706989	16	78569957	intron	WWOX	1.00 x10^{-06}	NHGRI	21107309
Diabetes Mellitus	rs414723	16	79228250	intron	WWOX	2.34 x10^{-05}	dbGaP	0
Bipolar Disorder	rs8047442	16	78764097	intron	WWOX	2.97 x10^{-05}	dbGaP	0
Schizophrenia	rs7200634	16	79176337	intron	WWOX	7.11 x10^{-05}	dbGaP	0
Alcohol Drinking	rs10515153	5	71989150	intergenic	ZNF366	3.50 x10^{-15}	dbGaP	0

dbGaP database of Genotypes and Phenotypes, *NHGRI* National Human Genome Research Institute; Data shown are compiled from both NHGRI GWAS Catalog (Source = NHGRI) and from most significant hits across analyses submitted to dbGaP. Top table: any trait with p value < 5x10^{-8}, bottom table any depression related trait with p value < 1x10^{-5}.

Supplementary Table 5 Top 25 hits from gene-based analysis of GWIS meta-analysis

SYMBOL	CHR	START	STOP	NSNPS	*p* value
ZNF366	5	71728479	71813554	22	$1.48x10^{-7}$
PHF3	6	64335725	64499229	13	$1.02x10^{-5}$
ANO4	12	101101304	101532419	142	$1.81x10^{-5}$
ABCC10	6	43385104	43428168	15	$1.02x10^{-4}$
LATS2	13	21537171	21645686	24	$1.04x10^{-4}$
ABCA2	9	139891686	139933367	4	$1.06x10^{-4}$
PRKG1	10	52740945	54068110	378	$1.27x10^{-4}$
PPARA	22	46536424	46649653	40	$1.50x10^{-4}$
C9orf139	9	139911916	139941234	8	$1.73x10^{-4}$
AMBRA1	11	46407964	46625675	10	$1.83x10^{-4}$
ZFYVE28	4	2261309	2430390	36	$2.76x10^{-4}$
C4orf22	4	81246874	81894910	54	$2.86x10^{-4}$
ATG13	11	46628826	46706368	9	$2.9x10^{-4}$
CBR4	4	169774921	169941426	20	$3.42x10^{-4}$
MDK	11	46392306	46415375	1	$3.6x10^{-4}$
ABCG1	21	43609799	43727354	77	$5.49x10^{-4}$
F2	11	46730730	46771056	5	$5.57x10^{-4}$
PRKCD	3	53180025	53236733	10	$6.54x10^{-4}$
SMARCA4	19	11061598	11186071	17	$6.76x10^{-4}$
PTPN3	9	112127746	112270590	34	$6.98x10^{-4}$
FUT7	9	139914626	139937462	6	$7.36x10^{-4}$
DGKZ	11	46344455	46412104	3	$7.48x10^{-4}$
MYOM3	1	24372525	24448665	37	$7.8x10^{-4}$
LCE3D	1	152541857	152562980	4	$7.9x10^{-4}$
LIMK2	22	31598225	31686066	9	$8.09x10^{-4}$

START/STOP the annotation boundaries of the gene including 10kb window on each side. *NSNPS* the number of SNPs annotated to that gene that were found in the meta-analysis summary statistics.

Supplementary Table 6 Summary results from polygenic risk score (PRS) analysis using PRSice-2

MDD risk prediction in UK Biobank							
Summary Statistics	**Threshold**	**R^2**	***p***	**β coefficient**	**S.E.**	**# SNP**	**Empirical-*p***
GS GWIS	0.001	1.49×10^{-5}	0.549664	0.00175134	0.00292733	345	-
GS GWIS	0.005	5.99×10^{-5}	0.230563	0.00348988	0.0029108	1526	0.684132
GS GWIS	0.01	4.22×10^{-6}	0.750409	0.000927105	0.00291447	2894	-
GS GWIS	0.02	3.24×10^{-6}	0.780509	-0.00081019	0.00290746	5363	-
GS GWIS	0.03	7.06×10^{-6}	0.680748	-0.00119544	0.00290545	7616	-
GS GWIS	0.04	1.23×10^{-6}	0.863456	0.000498521	0.00289875	9795	-
GS GWIS	0.05	1.12×10^{-6}	0.869843	0.00047545	0.00290157	11829	-
GS GWIS	0.1	3.75×10^{-6}	0.764254	-0.000873876	0.00291387	21270	-
GS GWIS	0.2	8.86×10^{-6}	0.644935	-0.00133834	0.00290428	37187	-
GS GWIS	0.3	7.40×10^{-6}	0.673605	-0.00122346	0.00290462	50757	-
GS GWIS	0.4	1.36×10^{-5}	0.567393	-0.00166105	0.00290442	62367	-
GS GWIS	0.5	3.15×10^{-5}	0.384791	-0.00252527	0.00290555	72518	-
GS GWIS	1	2.99×10^{-5}	0.397284	-0.00246053	0.00290671	105965	-
GS GWAS MDD	0.001	8.88×10^{-5}	0.144516	0.00423867	0.00290473	321	-
GS GWAS MDD	0.005	0.000678024	5.51×10^{-5}	0.0116971	0.00289997	1551	-
GS GWAS MDD	0.01	0.000509482	0.000472948	0.0100827	0.00288396	2950	-
GS GWAS MDD	0.02	0.000877145	4.49×10^{-6}	0.0133017	0.00289912	5421	-
GS GWAS MDD	**0.03**	**0.000959139**	**1.61×10^{-6}**	**0.0140196**	**0.00292193**	**7725**	**0.0001**
GS GWAS MDD	0.04	0.000931599	2.27×10^{-6}	0.0138323	0.00292524	9972	-
GS GWAS MDD	0.05	0.000752813	2.14×10^{-5}	0.0124522	0.00292971	12107	-
GS GWAS MDD	0.1	0.000718807	3.29×10^{-5}	0.0121979	0.00293704	21666	-

GS GWAS MDD	0.2	0.000558608	0.000251893	0.0107481	0.00293592	37657	-
GS GWAS MDD	0.3	0.000533753	0.000346304	0.0105007	0.00293441	50970	-
GS GWAS MDD	0.4	0.00055274	0.000271534	0.0106857	0.00293433	62381	-
GS GWAS MDD	0.5	0.000509722	0.00047149	0.0102732	0.00293775	72553	-
GS GWAS MDD	1	0.000451602	0.000998201	0.0096645	0.00293624	105850	-
GS GWAS EPQN	0.001	0.000379194	0.00256452	0.00877471	0.00290943	397	-
GS GWAS EPQN	0.005	0.00067407	5.80×10^{-5}	0.0117385	0.00291877	1689	-
GS GWAS EPQN	0.01	0.000481373	0.000679289	0.00991287	0.00291704	3049	-
GS GWAS EPQN	0.02	0.000669562	6.14×10^{-5}	0.0116189	0.00289876	5651	-
GS GWAS EPQN	0.03	0.000623175	0.000110568	0.0112464	0.00290845	7957	-
GS GWAS EPQN	0.04	0.000739668	2.53×10^{-5}	0.0122976	0.00291895	10159	-
GS GWAS EPQN	**0.05**	**0.000774376**	**1.63×10^{-5}**	**0.0125928**	**0.00292121**	**12296**	**0.0005**
GS GWAS EPQN	0.1	0.00062769	0.000104398	0.0112937	0.00291014	21790	-
GS GWAS EPQN	0.2	0.000755408	2.07×10^{-5}	0.0124031	0.00291313	37772	-
GS GWAS EPQN	0.3	0.000483159	0.000663811	0.00991244	0.00291151	51115	-
GS GWAS EPQN	0.4	0.000604933	0.000139459	0.0110822	0.00290889	62611	-
GS GWAS EPQN	0.5	0.000600135	0.00014825	0.0110556	0.0029135	72724	-
GS GWAS EPQN	1	0.00056519	0.00023156	0.0107335	0.00291479	106022	-
PGC2 GWAS MDD	0.001	0.00240871	2.86×10^{-14}	0.0221036	0.00290481	692	-
PGC2 GWAS MDD	0.005	0.00367058	5.98×10^{-21}	0.0273938	0.00291439	2259	-
PGC2 GWAS MDD	0.01	0.00414411	1.88×10^{-23}	0.0290898	0.00291192	3818	-
PGC2 GWAS MDD	0.02	0.00480299	6.26×10^{-27}	0.0312117	0.00290113	6501	-
PGC2 GWAS MDD	0.03	0.00548407	1.60×10^{-30}	0.0333745	0.00290212	8830	-
PGC2 GWAS MDD	0.04	0.00618157	3.34×10^{-34}	0.0353636	0.00289534	10975	-
PGC2 GWAS MDD	0.05	0.00598545	3.61×10^{-33}	0.0347968	0.00289553	12975	-
PGC2 GWAS MDD	0.1	0.0063685	3.45×10^{-35}	0.0360319	0.00290615	21544	-

PGC2 GWAS MDD	0.2	0.00634524	4.57×10^{-35}	0.0360177	0.00291037	35023	-
PGC2 GWAS MDD	0.3	0.00655498	3.58×10^{-36}	0.0366449	0.00291297	46240	-
PGC2 GWAS MDD	0.4	0.00659633	2.17×10^{-36}	0.0367308	0.00291056	55832	-
PGC2 GWAS MDD	**0.5**	**0.00674968**	**3.36×10^{-37}**	**0.0371502**	**0.00290994**	**64113**	**0.0001**
PGC2 GWAS MDD	1	0.00659775	2.13×10^{-36}	0.0367369	0.00291073	90514	-
GPC GWAS EPQN	0.001	5.93×10^{-5}	0.232975	0.0034842	0.00292112	450	-
GPC GWAS EPQN	0.005	0.00047754	0.000713734	0.00984361	0.00290826	1923	-
GPC GWAS EPQN	0.01	0.000349011	0.0038147	0.00846158	0.00292446	3508	-
GPC GWAS EPQN	0.02	0.000389339	0.00224533	0.00896142	0.00293236	6310	-
GPC GWAS EPQN	**0.03**	**0.000484233**	**0.000654681**	**0.00996823**	**0.00292465**	**8761**	**0.005999**
GPC GWAS EPQN	0.04	0.000389933	0.00222793	0.00895165	0.00292693	11106	-
GPC GWAS EPQN	0.05	0.000431876	0.00128937	0.00942108	0.00292695	13319	-
GPC GWAS EPQN	0.1	0.000412993	0.00164856	0.00918914	0.00291946	23043	-
GPC GWAS EPQN	0.2	0.000404552	0.00184043	0.00905106	0.00290545	38773	-
GPC GWAS EPQN	0.3	0.000337535	0.00443948	0.00827244	0.0029073	51732	-
GPC GWAS EPQN	0.4	0.000309518	0.00644059	0.00793176	0.00291105	62532	-
GPC GWAS EPQN	0.5	0.000288657	0.00851247	0.00764975	0.00290725	71717	-
GPC GWAS EPQN	1	0.000335342	0.0045702	0.00824335	0.00290654	100702	-

MDD risk prediction in Generation Scotland

Summary Statistics	Threshold	R^2	*p*	β coefficient	S.E.	# SNP	Empirical-*p*
UKB GWIS	0.001	0.000615863	0.0641373	0.0514063	0.0277688	359	-
UKB GWIS	**0.005**	**0.00140702**	**0.00515954**	**0.0777816**	**0.0278099**	**1626**	**0.039896**
UKB GWIS	0.01	0.00134793	0.00617577	0.0762639	0.0278508	3093	-
UKB GWIS	0.02	0.00082812	0.0318182	0.0597164	0.0278179	5620	-
UKB GWIS	0.03	0.000783734	0.0367345	0.0582806	0.0279028	7910	-

UKB GWIS	0.04	0.000783197	0.0367997	0.0583116	0.0279273	10172	-
UKB GWIS	0.05	0.000634099	0.0603104	0.05248	0.0279369	12296	-
UKB GWIS	0.1	0.000298689	0.197258	0.0361067	0.0280027	21828	-
UKB GWIS	0.2	5.56×10^{-5}	0.5779	0.0156126	0.0280572	38016	-
UKB GWIS	0.3	1.47×10^{-5}	0.774716	0.00803806	0.0280843	51652	-
UKB GWIS	0.4	8.31×10^{-6}	0.829698	0.00603532	0.0280596	63278	-
UKB GWIS	0.5	1.24×10^{-6}	0.933827	0.00233242	0.028091	73445	-
UKB GWIS	1	1.30×10^{-8}	0.993204	-0.000239447	0.0281123	107168	-
UKB GWAS MDD	0.001	0.000178845	0.318403	0.0331873	0.0332622	459	-
UKB GWAS MDD	0.005	0.000806406	0.0341973	0.0659184	0.0311268	1819	-
UKB GWAS MDD	0.01	0.000404563	0.133579	0.0461141	0.0307399	3293	-
UKB GWAS MDD	0.02	0.000888844	0.0262094	0.0662525	0.0298019	6055	-
UKB GWAS MDD	0.03	0.00157766	0.00306888	0.088043	0.0297367	8577	-
UKB GWAS MDD	0.04	0.00181354	0.00150359	0.0936559	0.0295073	10795	-
UKB GWAS MDD	0.05	0.00168391	0.00222044	0.0898185	0.0293616	12997	-
UKB GWAS MDD	**0.1**	**0.00265037**	**0.000125175**	**0.111487**	**0.029065**	**22771**	**0.0015**
UKB GWAS MDD	0.2	0.00243232	0.000238189	0.105894	0.0288175	38933	-
UKB GWAS MDD	0.3	0.00205055	0.000740482	0.0968134	0.0286928	52382	-
UKB GWAS MDD	0.4	0.00188791	0.00120425	0.0928607	0.0286796	63788	-
UKB GWAS MDD	0.5	0.00188513	0.00121414	0.092669	0.028641	73870	-
UKB GWAS MDD	1	0.0018228	0.00146334	0.0909943	0.0285978	107102	-
UKB GWAS EPQN	0.001	0.00257189	0.000157057	0.153412	0.0405891	937	-
UKB GWAS EPQN	0.005	0.00165949	0.00238208	0.111332	0.0366475	2886	-
UKB GWAS EPQN	0.01	0.00162997	0.00261083	0.104237	0.034628	4650	-
UKB GWAS EPQN	0.02	0.00192175	0.00108078	0.108054	0.0330582	7726	-
UKB GWAS EPQN	0.03	0.00222809	0.000434305	0.114407	0.0325177	10453	-

UKB GWAS EPQN	0.04	0.00260106	0.000144048	0.121308	0.0319139	12922	-
UKB GWAS EPQN	0.05	0.00300612	4.39×10^{-5}	0.129371	0.0316639	15217	-
UKB GWAS EPQN	0.1	0.00298832	4.62×10^{-5}	0.125519	0.0308111	25209	-
UKB GWAS EPQN	0.2	0.00291587	5.70×10^{-5}	0.121565	0.0302027	41077	-
UKB GWAS EPQN	0.3	0.00322564	2.31×10^{-5}	0.126745	0.0299438	54140	-
UKB GWAS EPQN	**0.4**	**0.00330371**	**1.84×10^{-5}**	**0.128011**	**0.029884**	**65276**	**0.0002**
UKB GWAS EPQN	0.5	0.00322111	2.34×10^{-5}	0.126075	0.029805	74888	-
UKB GWAS EPQN	1	0.0031251	3.09×10^{-5}	0.124076	0.0297791	106865	-
PGC2 GWAS MDD	0.001	0.00414855	1.62×10^{-6}	0.216448	0.0451304	770	-
PGC2 GWAS MDD	0.005	0.00521061	7.66×10^{-8}	0.203259	0.0378157	2458	-
PGC2 GWAS MDD	0.01	0.00560683	2.47×10^{-8}	0.196584	0.0352614	4098	-
PGC2 GWAS MDD	0.02	0.00596288	9.04×10^{-9}	0.195246	0.0339684	6870	-
PGC2 GWAS MDD	0.03	0.00557897	2.67×10^{-8}	0.18491	0.0332474	9285	-
PGC2 GWAS MDD	0.04	0.00548094	3.56×10^{-8}	0.179858	0.0326349	11584	-
PGC2 GWAS MDD	0.05	0.00656878	1.63×10^{-9}	0.193818	0.0321375	13556	-
PGC2 GWAS MDD	0.1	0.00774814	6.02×10^{-11}	0.205598	0.0314216	22338	-
PGC2 GWAS MDD	0.2	0.00887224	2.59×10^{-12}	0.216279	0.0309054	36238	-
PGC2 GWAS MDD	0.3	0.00939702	6.08×10^{-13}	0.221812	0.030813	47465	-
PGC2 GWAS MDD	0.4	0.00968321	2.74×10^{-13}	0.224167	0.0306811	57007	-
PGC2 GWAS MDD	0.5	0.00914253	1.23×10^{-12}	0.217315	0.0305994	65574	-
PGC2 GWAS MDD	**1**	**0.00992732**	**1.39×10^{-13}**	**0.22645**	**0.0306121**	**92248**	**0.0001**
GPC GWAS EPQN	0.001	0.000771689	0.0382666	0.0579815	0.0279835	453	-
GPC GWAS EPQN	0.005	0.000879223	0.0269669	0.0617304	0.0279071	1933	-
GPC GWAS EPQN	0.01	0.00107656	0.0143831	0.0685033	0.0279885	3521	0.10379
GPC GWAS EPQN	0.02	0.000736456	0.0429258	0.0568248	0.0280695	6348	-

GPC GWAS EPQN	0.03	0.000494566	0.0971083	0.0463769	0.0279541	8807	-
GPC GWAS EPQN	0.04	0.000815662	0.0331399	0.0594976	0.0279281	11165	-
GPC GWAS EPQN	0.05	0.000549567	0.0803194	0.0488544	0.0279354	13399	-
GPC GWAS EPQN	0.1	0.000292343	0.202061	0.0356931	0.0279792	23180	-
GPC GWAS EPQN	0.2	0.000389567	0.140913	0.04123	0.0280019	39079	-
GPC GWAS EPQN	0.3	0.00031477	0.185646	0.0370725	0.0280095	52223	-
GPC GWAS EPQN	0.4	0.000457649	0.110546	0.0446962	0.0280096	63160	-
GPC GWAS EPQN	0.5	0.000539128	0.0833026	0.0485336	0.0280243	72494	-
GPC GWAS EPQN	1	0.000611543	0.0651206	0.0516962	0.0280283	101954	-

In red, best-fit PRS. In **bold**, significant PRS after 10,000 permutations (Empirical-p < 0.05).

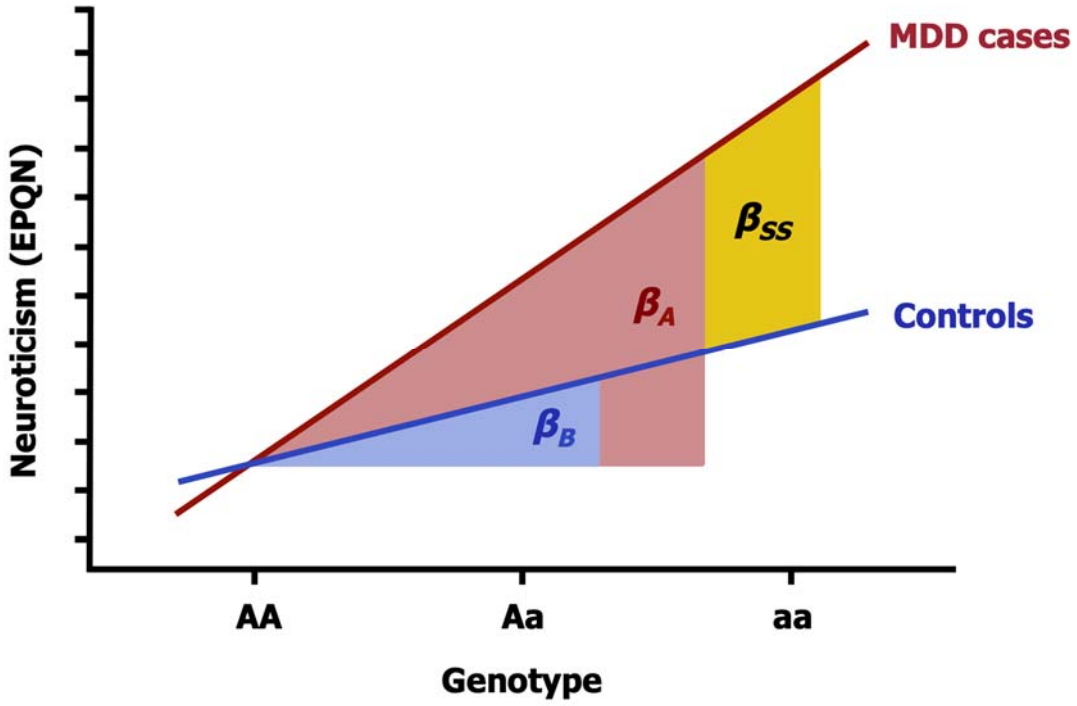
Supplementary Table 7 MDD stratification

		MDD quintiles			Stress-sensitivity quintile			MDD*Stress-sensitivity quintiles interaction		
Stratifier	**Sample size**	**effect**	**SE**	***p* value**	**effect**	**SE**	***p* value**	**effect**	**SE**	***p* value**
SPQ	1 093	0.181	0.223	0.416	0.189	0.218	0.387	-0.054	0.067	0.419
MDQ	1 022	-0.079	0.179	0.659	0.002	0.177	0.99	0.037	0.054	0.496
EPQN	2 010	0.265	0.121	**0.028**	0.09	0.12	0.453	-0.051	0.036	0.162
MDD course	2 016	0.044	0.275	0.875	0.013	0.275	0.964	-0.029	0.083	0.726
Age of onset	1 964	-0.419	0.375	0.264	-0.572	0.377	0.129	0.127	0.114	0.264
Episode count	2 016	-0.332	0.381	0.384	-0.258	0.379	0.496	0.121	0.115	0.295
Sex	2 016	0.008	0.017	0.626	0.01	0.017	0.545	-0.005	0.005	0.33

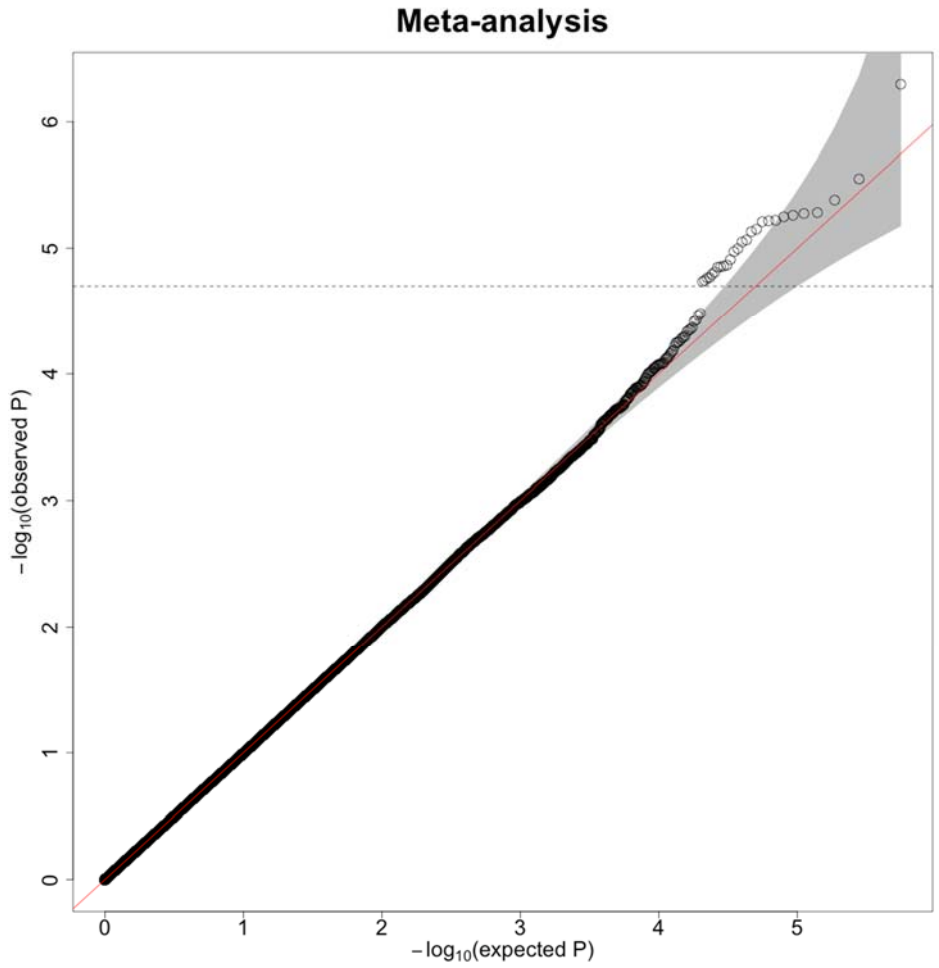
SPQ Schizotypal Personality Questionnaire score, MDQ Mood Disorder Questionnaire score, EPQN Eysenck Personality Questionnaire-Revised Short Form's Neuroticism Scale. Model: Stratifier ~ MDD quintile + Stress-sensitivity quintile + MDD*SS quintiles interaction + Age + Sex.

Stratifier	**mean**	**median**	**standard deviation**	**variance**
SPQ	5.90	5	4.47	19.97
MDQ	3.85	3	3.56	12.68
EPQN	6.42	7	3.32	10.99
MDD course	1.49	1	0.50	0.25
Age of onset	32.59	32	12.40	153.65
Episode count	6.19	1	10.45	109.26
Sex	1.67	2	0.47	0.22

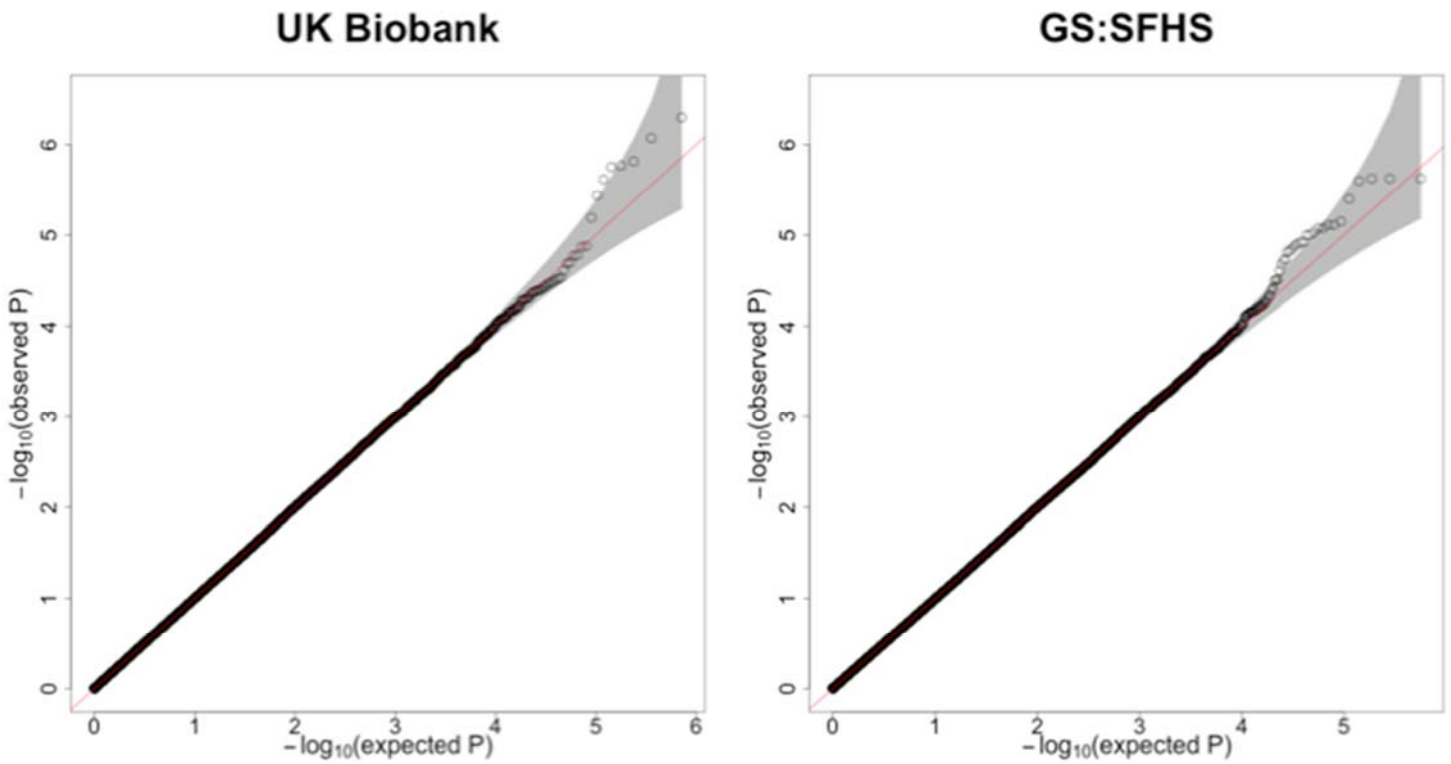
A.5 Supplementary Figures



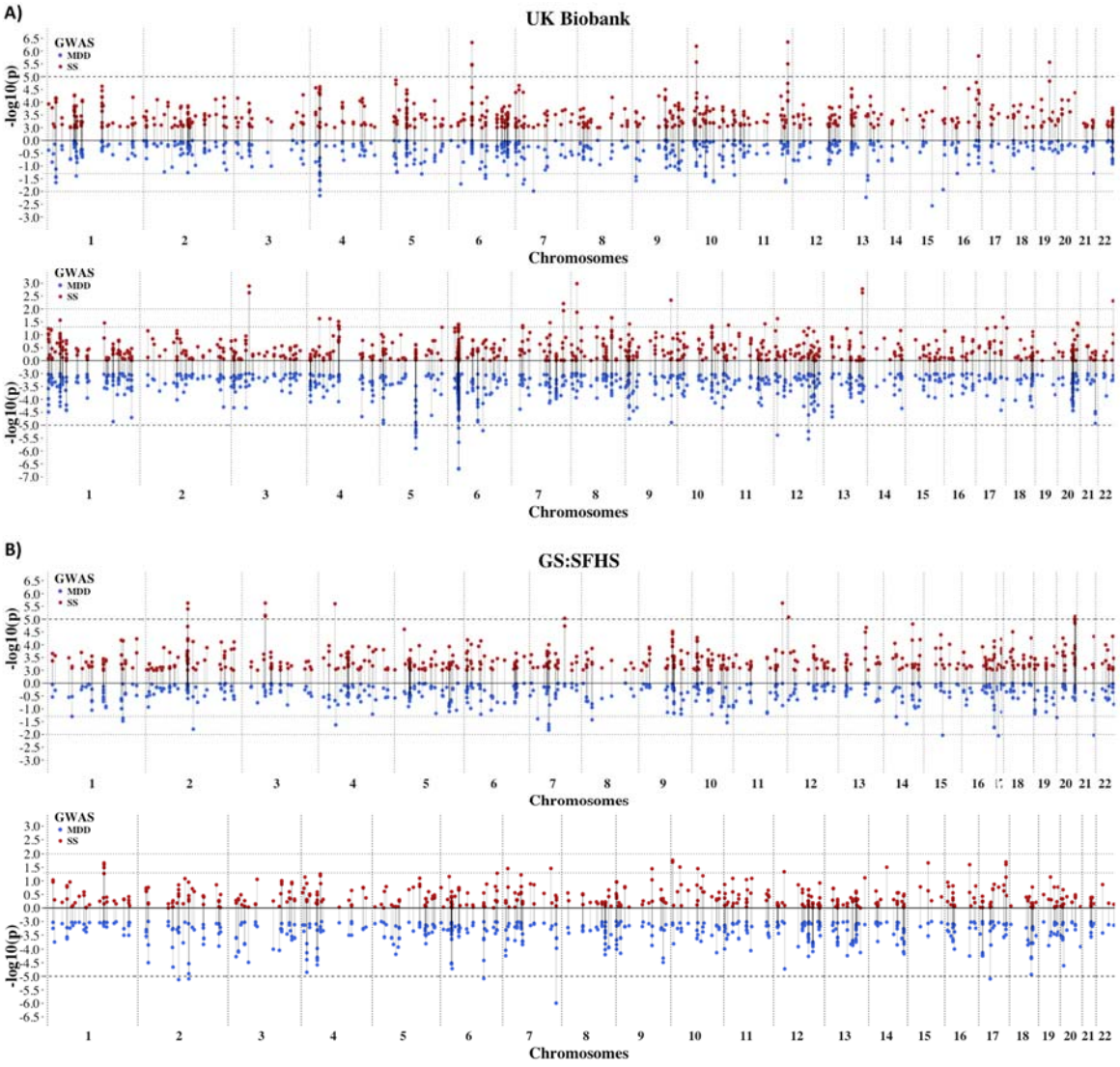
Supplementary Figure 1 Genetic stress-sensitivity effect representation. Genetic stress-sensitivity effect on MDD (β_{SS}) is defined as the difference between the regression coefficient in MDD cases (β_A) and the regression coefficient in controls (β_B) from linear models regressed on EPQN, adjusted by covariates. X-axis represents the SNP genotype as homozygous for the major allele (AA), heterozygous (Aa) and homozygous for the minor allele (aa).



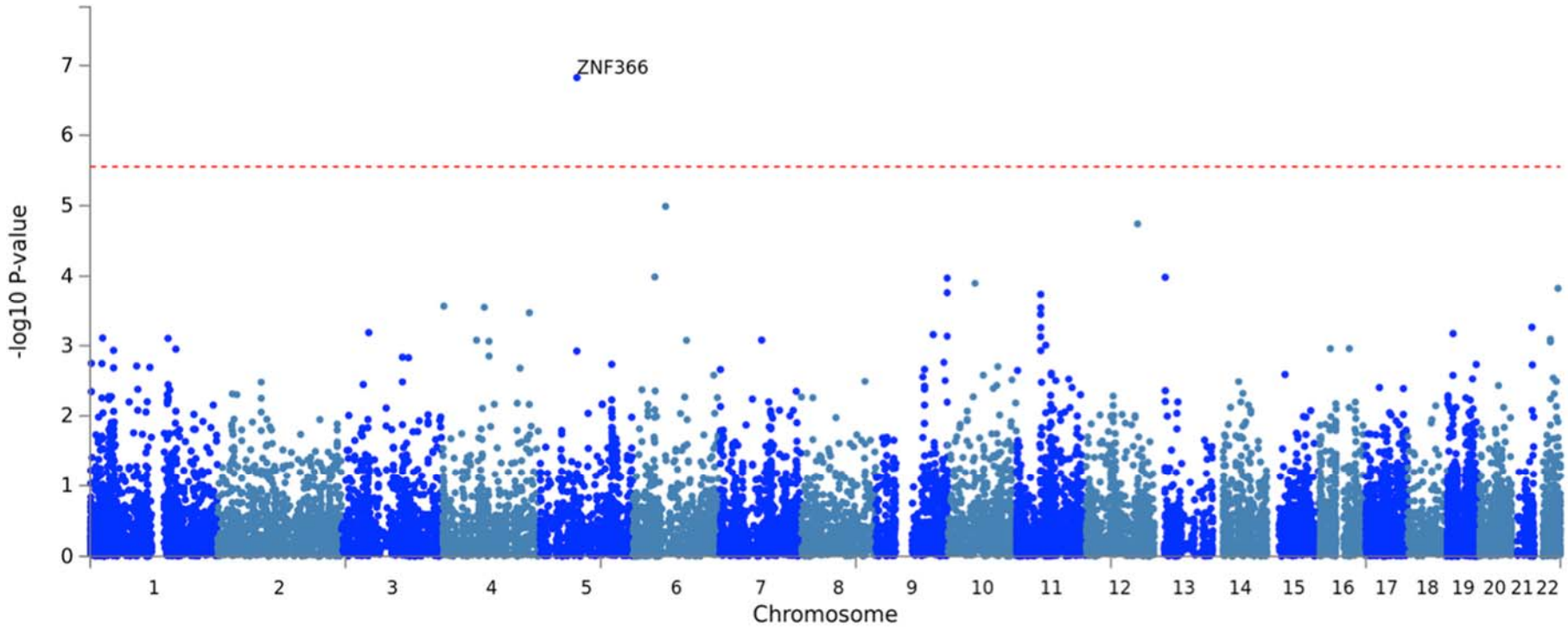
Supplementary Figure 2 QQ plot from stress-sensitivity meta-analysis. QQ plot of GWIS from sample size weighted meta-analysis (λ = 0.997; s.e. = 1.05 x 10^{-5}). All SNPs wit $p < 2 \times 10^{-5}$, *p*-threshold (dot line) where some SNPs start to deviate from null distribution going outside 95% confidence intervals (grey shadow), were selected to perform DEPICT analyses to assess pathway and functional genomic analyses. 27 top variants from 12 independent loci were selected.



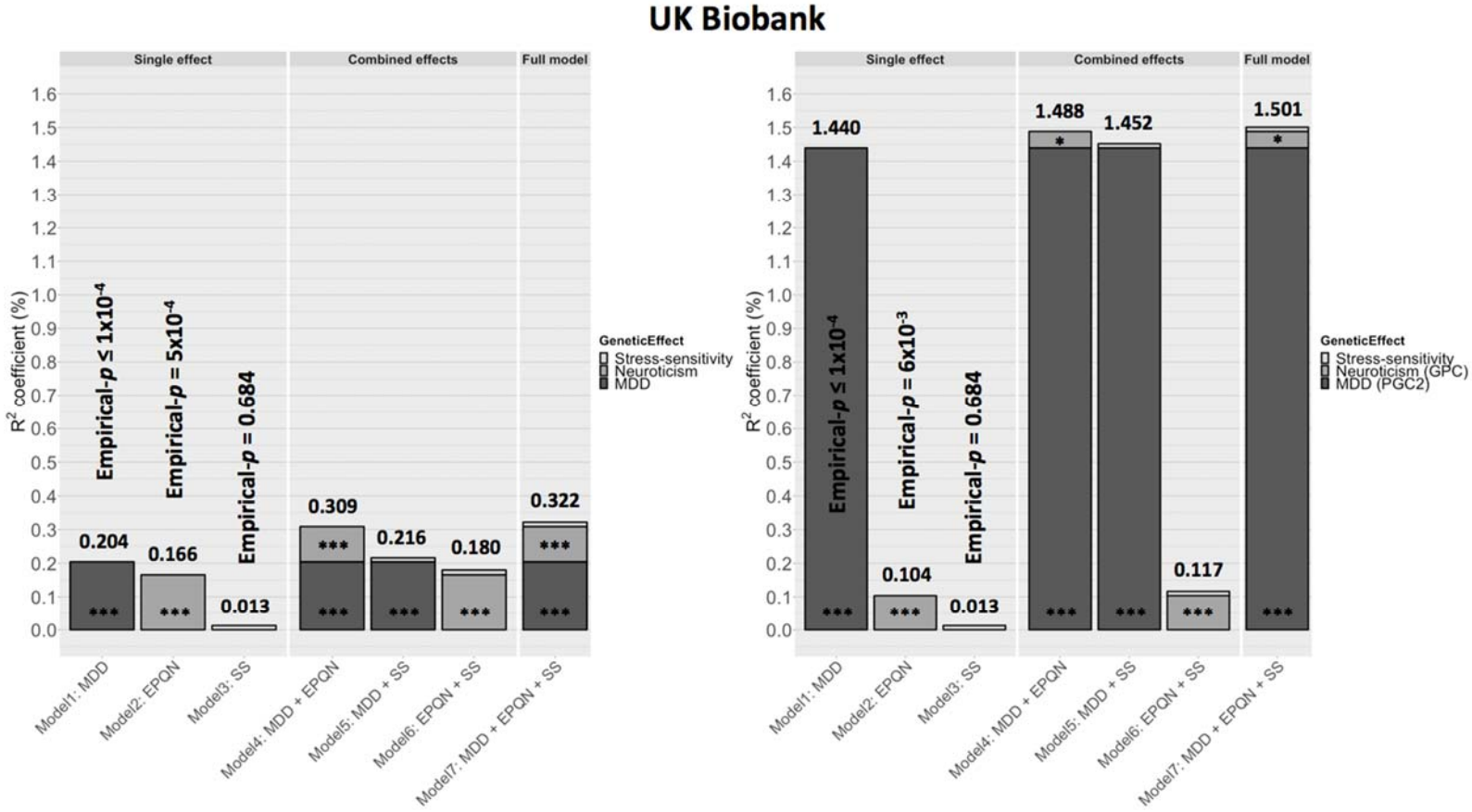
Supplementary Figure 3 QQ plots of GWIS *p* values. QQ plots of GWIS from (A) UKB (λ = 1.014; s.e.= 1.027 x 10^{-5}), (B) GS:SFHS (λ = 0.997; s.e. = 7.989 x 10^{-6}). The 95% confidence interval is shaded in grey.



Supplementary Figure 4 Miami plots on UK Biobank and Generation Scotland: Scottish Family Health Study. Miami plots showing comparison between association profile between SS and MDD main additive effects. Miami plots from (A) UKB filtering for SS *p* values (top) and MDD *p* values (bottom), (B) GS:SFHS filtering for SS *p* values (top) and MDD *p* values (bottom). Filter at $p = 1 \times 10^{-3}$. The x-axis is base-paired chromosomal position and y-axis is the significance (-log10 *p*) of association with (up; red dots) SS effect and (down; blue dots) MDD. Dot line: genome-wide suggestive threshold ($p = 1 \times 10^{-5}$) at the filtered effect; dashes lines: *p* value = 0.01 and 0.05 at compared effect.



Supplementary Figure 5 Manhattan plot of the gene-based test for stress-sensitivity. Manhattan plot showing gene-based association of stress-sensitivity. The x-axis is base-paired chromosomal position and y-axis is the significance ($-\log_{10} p$ value) of association with SS effect. Genome-wide significance threshold showed by red dashed line was defined at $p = 0.05/17{,}931 = 2.79 \times 10^{-6}$.



Supplementary Figure 6 PRS profiling predicting MDD in UK Biobank. MDD risk explained (R^2 coefficient (%); top bar values) on the liability scale by each PRS in UKB; weighted by GWAS main additive and GWIS stress-sensitivity effects independently and combined. (A) Using summary statistics from GS:SFHS as discovery sample. (B) Replication fitting PRS_D and PRS_N using summary statistics from worldwide consortiums (i.e. PGC & GPC). Significance codes: *p* values *** < 0.001 < ** < 0.01 < * < 0.05; derived from likelihood ratio tests. *SS* stands for stress-sensitivity.

A.6 Arnau-Soler *et al.*, 2018, PLOS ONE

PLOS ONE

RESEARCH ARTICLE

Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder

Aleix Arnau-Soler[1]*, Mark J. Adams[2], Generation Scotland[¶], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[¶], Caroline Hayward[3], Pippa A. Thomson[1]*

1 Medical Genetics Section, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **2** Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, United Kingdom, **3** Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

¶ Generation Scotland is a collaboration between the University Medical School and NHS in Aberdeen, Dundee, Edinburgh and Glasgow, Scotland, United Kingdom. Membership list is provided in the Acknowledgments. Membership of the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium is provided in the Acknowledgments.
* Pippa.Thomson@ed.ac.uk (PT); aleix.arnau.soler@igmm.ed.ac.uk (AAS)

Citation: Arnau-Soler A, Adams MJ, Generation Scotland, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Hayward C, Thomson PA (2018) Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. PLoS ONE 13(12): e0209160. https://doi.org/10.1371/journal.pone.0209160

Editor: Zhenghui Yi, Shanghai Mental Health Center, CHINA

Received: September 3, 2018

Accepted: December 2, 2018

Published: December 20, 2018

Data Availability Statement: Summary statistics generated (i.e. GWIS from Generation Scotland, GWIS from UK Biobank, and meta-analysed GWIS have been added as supplementary data and uploaded in bioRxiv (DOI: 10.1101/194290). This data is also available under requests at the Institute of Genetics and Molecular Medicine (IGMM) by contacting Dr. Pippa Thomson (pippa.thomson@ed.ac.uk). Data from the samples used in this study (Generation Scotland and UK Biobank)

Abstract

Individual response to stress is correlated with neuroticism and is an important predictor of both neuroticism and the onset of major depressive disorder (MDD). Identification of the genetics underpinning individual differences in response to negative events (stress-sensitivity) may improve our understanding of the molecular pathways involved, and its association with stress-related illnesses. We sought to generate a proxy for stress-sensitivity through modelling the interaction between SNP allele and MDD status on neuroticism score in order to identify genetic variants that contribute to the higher neuroticism seen in individuals with a lifetime diagnosis of depression compared to unaffected individuals. Meta-analysis of genome-wide interaction studies (GWIS) in UK Biobank (N = 23,092) and Generation Scotland: Scottish Family Health Study (N = 7,155) identified no genome-wide significance SNP interactions. However, gene-based tests identified a genome-wide significant gene, *ZNF366*, a negative regulator of glucocorticoid receptor function implicated in alcohol dependence ($p = 1.48 \times 10^{-7}$; Bonferroni-corrected significance threshold $p < 2.79 \times 10^{-6}$). Using summary statistics from the stress-sensitivity term of the GWIS, SNP heritability for stress-sensitivity was estimated at 5.0%. In models fitting polygenic risk scores of both MDD and neuroticism derived from independent GWAS, we show that polygenic risk scores derived from the UK Biobank stress-sensitivity GWIS significantly improved the prediction of MDD in Generation Scotland. This study may improve interpretation of larger genome-wide association studies of MDD and other stress-related illnesses, and the understanding of the etiological mechanisms underpinning stress-sensitivity.

is third-party data that comes from consortiums with their own Data Access Committee. Data is available from the MRC IGMM Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data. Generation Scotland data is available to researchers on application to the Generation Scotland Access Committee (access@generationscotland.org). UK Biobank data is available to researchers on application via the UK Biobank Resource Access Management System (access@ukbiobank.ac.uk). The managed access a process ensures that approval is granted only to research that comes under the terms of participant consent.

Funding: Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL), Reference 104036/Z/14/Z). The Psychiatric Genomics Consortium has received major funding from the US National Institute of Mental Health and the US National Institute of Drug Abuse (U01 MH109528 and U01 MH1095320).The 1st author AAS is funded by University of Edinburgh (www.ed.ac.uk) and Medical Research Council for his PhD study at the University of Edinburgh Institute of Genetics and Molecular Medicine (www.ed.ac.uk/igmm). MA is supported by the Wellcome Trust Strategic Award STRADL (Reference 104036/Z/14/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium" was involved in this study. The following companies are affiliated to this Consortium: Humus, Reykjavik, IS, Solid Biosciences, Boston, deCODE Genetics / Amgen, Reykjavik, Roche, Janssen and Pfizer. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

Introduction

Stressful life events are known to increase liability to mental illness and disease-related traits [1] including neuroticism [2–4], major depressive disorder (MDD) [5–7], autoimmune diseases [8] and some cancers [9, 10]. A greater understanding of the causal mechanism by which negative events affect disease risk or outcome may be beneficial in identifying individuals for targeted support. However, it has been proposed that sensitivity to stress may be an important predictor of response to stress [11, 12]. In particular, the effect on an individual may result more from the perceived stress than the event itself, and may be dependent on individual differences in stress-sensitivity [13–18]. Studies of 5-HTT and twin studies suggest that stress-sensitivity may, at least in part, be heritable [19–22]. Despite a complex interaction between MDD, neuroticism and stress, multivariate structural equation models have confirmed a genetic effect on perceived stress, overlapping that on MDD or neuroticism, but with a specific genetic component [21]. The inter-relatedness of these traits may offer an approach to identify the genetic variation that affects an individual's stress-sensitivity, and improve genetic prediction of an individual's liability to negative outcomes. By modelling the interaction between SNP allele and MDD status on neuroticism score through genome-wide interaction studies (GWIS), we sought to investigate the genetics of stress-sensitivity.

The personality trait neuroticism is moderately heritable (30–50% estimates from twin studies) [23–26], is higher in individuals with depression compared to controls [27, 28] and is known to have shared genetic aetiology with depression [29–32]. Neuroticism is strongly correlated with measures of sensitivity to punishment but not reward [33], positively correlated with perceived personal relevance of a stressor [34, 35] and has been used previously as a proxy measure of stress-sensitivity [36]. Neuroticism is thought to mediate or interact with the effects of adverse life events on risk of depression [5, 37]. It has a substantial stable component [38], however, there is evidence for change, as well as stability, across the life span [2–4, 39]. Individual differences in neuroticism are enduringly influenced by both genetic and environmental factors [40]. Whereas the stable component of neuroticism is strongly determined by genetics, change in neuroticism score is attributed to the effects of unshared environment [39]. Persistent change in neuroticism score has been shown in response to life events [2–4]. Negative life events lead to small persistent increases in neuroticism over time [3]. However, recent stressful life events ($\beta = 0.14$ 95%CI 0.13–0.15, $p < 0.001$) have a stronger effect than distant stressful life events suggesting a reduction of effect over time [3]. Long-lasting increases in neuroticism associated with distant negative life events are mediated by depression [4].

Major depressive disorder (MDD) is a complex disorder influenced by both genetic contributions and environmental risk factors, with heritability estimates from twin and family studies of between 31–42% [41, 42]. Confirmed environmental risk factors for MDD include maternal infections, childhood maltreatment and negative life events [5–7, 43, 44]. However, few genetic studies have such information and even fewer prospective studies exist. Incorporation of stressful life events has been shown to improve the ability to predict MDD [45, 46] and, although stress is an environmental risk factor, it may have an independent genetic contribution to risk of depression [46–50].

These studies suggest that a genetic variable derived from the difference in neuroticism levels seen in individuals with MDD compared to controls may allow us to identify genetic loci important for stress-sensitivity. We sought to identify the genetic underpinnings of individual's sensitivity to stress response (stress-sensitivity) by identifying variants that contribute to the higher neuroticism levels seen in individuals with a lifetime diagnosis of MDD. Further, polygenic risk scores (PRS) derived from this stress-sensitivity variable may improve prediction of MDD over that based on MDD or neuroticism PRS alone.

Using unrelated individuals from two large population-based samples, UK Biobank (UKB; N = 23,092) and Generation Scotland: Scottish Family Health Study (GS:SFHS; N = 7,155), we sought to identify genes involved in stress-sensitivity by performing GWIS for the interaction between MDD status and SNP allele on neuroticism score. We identified a gene significantly associated with stress-sensitivity and show that a PRS derived from the interaction term of the GWIS, significantly predicts liability to depression independently of the PRS for MDD and/or neuroticism.

Materials and methods

UK Biobank (UKB) participants

UKB is a major national health resource that aims to improve the prevention, diagnosis and treatment of a wide range of illnesses. It recruited more than 500,000 participants aged from middle to older age who visited 22 assessment centres across the UK between 2006 and 2010. Data were collected on background and lifestyle, cognitive and physical assessments, sociodemographic factors and medical history. The scientific rationale, study design, ethical approval, survey methods, and limitations are reported elsewhere [51, 52]. UKB received ethical approval from the NHS National Research Ethics Service North West (Research Ethics Committee Reference Number: 11/NW/0382). All participants provided informed consent. The present study was conducted on genome-wide genotyping data available from the initial release of UKB data (released 2015). Details of sample processing specific to UKB project are available at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583 and the Axiom array at http://media.affymetrix.com/support/downloads/manuals/axiom_2_assay_auto_workflow_user_guide.pdf. UKB genotyping and the stringent QC protocol applied to UKB data before it was released can be found at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580. SNPs genotyped on GS:SFHS were extracted from the imputed UKB genotype data [53] (imputed by UKB using a merged panel of the UK10K haplotype reference panel and the 1000 Genomes Phase 3 reference panel) with quality > 0.9 was hard-called using PLINK v1.9 [54]. Individuals were removed based on UKB genomic analysis exclusion (UKB Data Dictionary item #22010), non-white British ancestry (#22006: genetic ethnic grouping; from those individuals who self-identified as British, principal component analysis was used to remove outliers), high genotype missingness (#22005), genetic relatedness (#22012; no pair of individuals have a KING-estimated kinship coefficient > 0.0442), QC failure in UK BiLEVE study (#22050 and #22051: UK BiLEVE Affymetrix and UK BiLEVE genotype quality controls for samples) and gender mismatch (#22001: genetic sex). Further, from the initial release of UKB data and using PLINK pi-hat < 0.05, individuals who were also participants of GS:SFHS and their relatives were excluded to remove any overlap of individuals between discovery and target samples. A dataset of 109,283 individuals with 557,813 SNPs remained for further analysis, aged 40–79 (57,328 female, 51,954 male; mean age = 57.1 years, s.d. = 7.99), of which 109,282 had data available for neuroticism score and 23,092 had data available on MDD status (n_{cases} = 7,834, $n_{controls}$ = 15,258, n_{female} = 11,510, n_{male} = 11,582; mean age = 57.7 years, s.d. = 8.04). Thus, the final dataset comprised 23,092 unrelated individuals.

Generation Scotland Scottish Family Health Study (GS:SFHS) participants

GS:SFHS is a family-based genetic epidemiology study which includes 23,960 participants from ~ 7,000 Scottish family groups collected by a cross-disciplinary collaboration of Scottish medical schools and the National Health Service (NHS) from Feb 2006 to Mar 2011. Participants were interviewed and clinically assessed for a wide range of health-related traits (including high-fidelity phenotyping for Major Depressive Disorder and related endophenotypes),

environmental covariates and linked to routine health records [55, 56]. All components of GS:SFHS obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the NHS (Research Ethics Committee Reference Number: 05/S1401/89) and participants provided written consent. The protocol for recruitment is described in detail in previous publications [57, 58]. GS:SFHS genotyping and quality control is detailed elsewhere [59]. Briefly, individuals with more than 2% missing genotypes and sex discrepancies were removed, as well as population outliers. SNPs with genotype missingness > 2%, minor allele frequency < 1% and a Hardy-Weinberg Equilibrium test $p < 1x10^{-6}$ were exclude. Finally, individuals were removed based on relatedness (pi-hat < 0.05), maximizing retention of case individuals, using PLINK v1.9 [54]. Genome-wide SNP data for further analysis comprised 7,233 unrelated individuals genotyped for 560,698 SNPs (n_{female} = 3,476, n_{male} = 3,757; PLINK v1.9 [54]), aged 18–92 (mean age = 50.4 years, s.d. = 12.06) of which: 7,190 had clinical data on MDD; 7,196 individuals had data on neuroticism; and 7,155 had data on both neuroticism and MDD.

Phenotype assessment

Neuroticism score (EPQN). Participants in both UKB and GS:SFHS cohorts were assessed for neuroticism using 12 questions from the Eysenck Personality Questionnaire-Revised Short Form's Neuroticism Scale (EPQN) [60–63]. Neuroticism can be scored by adding up the number of "Yes" responses on EPQN. This short scale has a reliability of more than 0.8 [64]. EPQN distributions were found to be sufficiently "normal" after assessment for skewness and kurtosis to be analysed using linear regression (both coefficients were between -1 and 1).

MDD diagnoses. In UKB, the MDD phenotype was derived following the definitions from Smith et al. [63] Current and previous depressive symptoms were assessed by items relating to the lifetime experience of minor and major depression [60], items from the Patient Health Questionnaire [65] and items on help-seeking for mental health [63]. Using a touchscreen questionnaire, participants were defined as probable cases if they i) answered "Yes" to the question "Ever depressed for a whole week" (UKB field: 4598), plus at least 2 weeks duration (UKB field: 4609), or ii) did report having seen a GP or psychiatrist for nerves, anxiety, tension or depression (UKB fields: 2090 and 2010) and reported symptoms (UKB field: 4631) with at least 2 weeks duration (UKB field: 5375). In our unrelated sample, 7,834 participants were diagnosed with MDD (with single, moderate or recurrent episodes) and 15,258 were controls (N = 23,092).

In GS:SFHS, participants took in-person clinical visits where they were screened for a history of psychiatric and emotional disorders (i.e., psychiatric, mood state/psychological distress, personality and cognitive assessment) by trained researchers using the Structured Clinical Interview for DSM-IV Non-Patient Version (SCID) [66], which is internationally validated to identify episodes of depression. Those participants that were positive in the initial screening continue through clinical interview and were administered the mood sections of the SCID. The SCID elicited the presence or absence of a lifetime history of MDD, age of onset and number of episodes. Participants fulfilling the criteria for at least one major depressive episode within the last month were defined as current MDD cases. Participants who were screened positive for Bipolar I Disorder were excluded. Those participants who were negative during the initial screening or did not fulfilled criteria for MDD were assigned as controls. Further details regarding the diagnostic assessment are reported elsewhere [56, 57]. All interviewers were trained for the administration of the SCID. Inter-rater reliability for the presence or absence of a lifetime diagnosis of major depressive disorder was good (Kappa = 0.86,

$p < 0.001$, 95%CI 0.7 to 1.0). In our unrelated GWIS sample (N = 7,155), 2,010 had a lifetime diagnosis of MDD and 5,145 were controls.

Statistical methods

GWIS and derivation of a genetic stress-sensitivity effect. The effect size of an stress-sensitivity effect (β_{SS}) was derived by performing a GWIS for the effect of the MDD status and SNP allele on EPQN (dependent variable) in both UKB and GS:SFHS cohorts using PLINK 1.90 (PLINK-command—gxe; fitting MDD diagnosis as a binary "group" effect) [54]. PLINK-command—gxe estimates the difference in allelic association with a quantitative trait (EPQN) between two groups (MDD cases vs. controls) producing effect estimates on each group and a test of significance for the interaction between SNP allele and MDD status. The interaction *p* value reflects the difference between the regression coefficient of the allelic effect in a linear model for EPQN in MDD cases (β_A) and the same regression coefficient in a linear model for EPQN in controls (β_B). The stress-sensitivity interaction effect was defined as the difference in allele effect between MDD cases and control groups.

Considering one SNP, the effect it confers to EPQN can be modelled by MDD status (control = 0, MDD case = 1) as follows:

$$\begin{cases} MDD = 0;\ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_A SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

This is equivalent to modelling the effect on MDD cases as follows:

$$\begin{cases} MDD = 0;\ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_B SNP + (\beta_A - \beta_B) SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

Or, it can be modelled as a whole as:

$$EPQN = \beta_0 + \beta_2 MDD + \beta_B SNP + (\beta_A - \beta_B) SNP * MDD + \beta_{0c} COV + \beta_{2C} COV * MDD + \varepsilon$$

Where *COV* stands for covariates, β_2 stands for $\beta_1 - \beta_0$, and β_{2c} stands for $\beta_{1c} - \beta_{0c}$.

Thus, the interaction effect (β_{SS}) can be estimated as the difference in allelic effect on EPQN between MDD cases (β_A) and controls (β_B) as follows,

$$\hat{\beta}_{SS} = \hat{\beta}_A - \hat{\beta}_B$$

$\hat{\beta}_{SS}$ is therefore defined as the effect size reflecting the genetic stress-sensitivity effect on MDD cases compared to controls (S1 Fig).

Stress-sensitivity GWIS, main additive effect GWASs, meta-analysis and gene-set analysis. For GWIS and subsequent analyses, sample specific covariates were applied as follows: *UKB*. All phenotypes were adjusted for centre, array and batch as random effects prior to analyses. Analyses were adjusted for age, sex and 15 informative principal components (PCs; UKB Data Dictionary items #22009.01 to #22009.15) as fixed effects to take account of possible population stratification. *GS:SFHS*. All the analyses were adjusted for age, sex and 20 PCs.

GWAS for MDD and neuroticism, using logistic and linear models of additive allelic effects respectively, were conducted on the same sample sets for comparison and generation of matched PRS using PRSice-2 [67].

Results from the GWIS of UKB and GS:SFHS were combined in a sample size weighted meta-analysis performed using METAL [68]. While the use of standard error weighting is more common, the different diagnostic scheme and MDD prevalence between the two cohorts

(GS:SFHS; 12.2%, UKB: 25.8%) [57, 63] may indicate systematic differences in the measurement of MDD. Generalized gene-based analysis of the meta-analysis was performed using MAGMA [69] implemented through FUMA [70] (http://fuma.ctglab.nl). Briefly, SNP summary statistics were mapped to 17,931 protein-coding genes. Individual SNP *p* values from a gene were combined into a gene test-statistic using a SNP-wise model and a known approximation of the sampling distribution used to obtain a gene-based *p* value. Genome-wide significance was defined at $p = 0.05/17{,}931 = 2.79\text{x}10^{-6}$.

LD Score regression. The summary statistics from the meta-analysis were used to examine the genetic overlap between the polygenic architecture of stress-sensitivity, MDD and neuroticism. LD score regression was used to derive the genetic correlations (r_G) between these traits [71, 72] using meta-analysed GWAS and GWIS summary statistics. SNP-based heritability was also estimated using LD score regression, using the summary statistics from single-SNP analyses.

Pathway, functional and gene expression analyses. Lead SNPs, independently associated with the phenotype, were identified using PLINK 1.90 by clumping (p threshold $< 2\text{x}10^{-5}$; LD r2 > 0.1; physical kb threshold = 500kb; 1000 Genomes Project Phase 1 CEU, GBR, TSI genotype data), and analysed using DEPICT [73]. Further detail is given in 'DEPICT analyses' in S1 Supporting Information.

Genes associated with lead SNPs were investigated for evidence of: phenotypic association in the NCBI dbGaP database of genotypes and phenotypes [74] (https://www.ncbi.nlm.nih.gov/gap/phegeni), regulatory DNA elements in normal cell lines and association with expression quantitative trait loci (eQTLs) using the RegulomeDB database [75] (http://www.regulomedb.org) and the Genotype-Tissue Expression (GTEx) Portal [76] (http://www.gtexportal.org).

Polygenic profiling. PRS were produced using PRSice-2 [67], permuted 10,000 times and standardized to a mean of 0 and a standard deviation of 1. Using GWIS summary statistics, we created PRS for stress-sensitivity (PRS_{SS}) by weighting the sum of the reference alleles in an individual by the stress-sensitivity effect (β_{SS}). Additional PRS were generated weighting by MDD main additive effects (PRS_D) and neuroticism main additive effects (PRS_N) using GWAS summary statistics from GS:SFHS or UKB. In addition, PRS_D and PRS_N were also generated using summary statistics from the most recent Psychiatric Genetic Consortium (PGC) MDD meta-analysis [42] (excluding GS:SFHS, and UKB individuals when required; N = 155,866 & 138,884) and the Genetics of Personality Consortium (GPC) neuroticism meta-analysis [24, 77] (N = 63,661). Generalized linear models were implemented in R 3.1.3 [78]. The direct effect of PRS_{SS} (model 1), PRS_D (model 2) and PRS_N (model 3) on MDD risk were assessed in independent logistic regression models on GS:SFHS (target cohort) using GWAS and GWIS statistics from UKB (the largest cohort) as the discovery sample to weight PRS. Multiple regression models fitting both PRS_D and PRS_N (model 4) and fitting each of them separately with PRS_{SS} (models 5 and 6) were also calculated. Finally, full additive multiple regression models fitting PRS weighted by all three effects (full model) was assessed using both PRS_{SS}, PRS_D and PRS_N at their best-fit in independent models. Further, results were also assessed using PRS_D and PRS_N weighted by PGC2 MDD [42] and GPC neuroticism [77] summary statistics. Further detail is given in 'Polygenic Profiling' in S1 Supporting Information. All models were adjusted by sex, age and 20 PCs. A null model was estimated from the direct effects of all covariates on MDD. 10,000 permutations were used to assess significance of each PRS. The predictive improvement of combining the effects of multiple PRS over a single PRS alone was tested for significance using the likelihood-ratio test.

Cross-validation was performed using UKB as target sample and GS:SFHS as discovery sample. Additional analyses using PRS_D and PRS_N weighted by PGC2 MDD [42] and GPC

neuroticism [77] summary statistics were also tested. MDD status on UKB was adjusted by centre, array and genotyping batch as random effects and scaled (between 0 and 1) prior to analysis, giving a quasi-binomial distribution of MDD status on UKB. Models implemented on UKB (quasi-binomial regression) were adjusted by sex, age and 15 PCs. Nagelkerke's R^2 coefficients were estimated to quantify the proportion of MDD liability explained at the observed scale by each model and converted into R^2 coefficients at the liability scale (prevalence: 12.2% in GS:SFHS [57] and 25.8% in UKB [63]) using Hong Lee's transformation [79] available from GEAR: GEnetic Analysis Repository [80].

Using stress-sensitivity to stratify depression. GS:SFHS MDD cases (n_{cases} = 2,016; n_{female} = 1,345, n_{male} = 671) have data available on MDD course (single or recurrent), age of onset (n = 1,964) and episode count (n = 2,016), as well as on neuroticism (n = 2,010). In addition, a subset were evaluated by Mood Disorder Questionnaire [81] (MDQ; n = 1,022) and Schizotypal Personality Questionnaire [82] (SPQ; n = 1,093). The reduced sample number of MDQ and SPQ reflects the later addition of these questionnaires to the study and does not reflect a particular subgroup of GS:SFHS.

Difference in PRS_{SS} and PRS_D between MDD cases and controls on GS:SFHS were tested using a Student's two sample t-test (two tailed). Cases of MDD on GS:SFHS with data available on each trait analyzed were stratified by quintiles based on PRS_{SS} and PRS_D (5x5 groups). Post hoc, the effects on each trait of quintiles based on PRS_{SS} and its interaction effect with quintiles based on PRS_D were assessed using linear regression models adjusting by sex and age in an attempt to identify a characteristic subtype of MDD patients with differential stress-sensitivity levels. The same analysis was reproduced using PRSs as continuous variables.

Results

We confirmed the elevated neuroticism score in MDD cases in our samples. Individuals with a diagnosis of MDD had significantly higher EPQN scores compared to healthy controls (all $p < 1.9\text{x}10^{-279}$) in both GS:SFHS ($mean_{controls}$ = 3.16; $mean_{cases}$ = 6.42) and UKB ($mean_{controls}$ = 2.79; $mean_{cases}$ = 5.64). Neuroticism levels differ significantly between males and females. To control for this and any age/polygenic effects, which may account for differences in the prevalence of MDD, we created a matched set of cases and controls. The difference in neuroticism levels between cases and controls remained significant after matching the controls for PGC PRS_D, sex and age. (GS:SFHS: $mean_{controls}$ = 3.51; UKB: $mean_{controls}$ = 2.97; all $p < 2.7\text{x}10^{-158}$; S1 Table).

Meta-analysis of stress-sensitivity in UKB and GS:SFHS

No SNPs were associated with stress-sensitivity at the genome-wide significant threshold ($p < 5\text{x}10^{-8}$, Fig 1). However, 14 SNPs from 8 loci achieved suggestive p value ($p < 1\text{x}10^{-5}$) ranging between $p = 8.9\text{x}10^{-6}$–$5.1\text{x}10^{-7}$ (summary statistics available in S1–S3 Files; Meta-analysis: Table 1; UKB and GS:SFHS: S2 and S3 Tables; Meta-analysis QQ-plot with λ: S2 Fig; UKB and GS:SFHS QQ-plots: S3 Fig). Traits with prior evidence of association with the nearest genes to the 8 lead SNPs were identified using dbGap and are shown in S4 Table. Comparison between the SNP association profile along the genome between stress-sensitivity GWIS and MDD GWAS meta-analyses is shown in Miami plots filtering for the most significant stress-sensitivity or MDD SNPs ($p < 0.001$; Meta-analysis: Fig 2; UKB and GS:SFHS: S4 Fig). No SNP with a p-value < 0.01 had a corresponding p-value in the alternate trait, suggesting that different variants contribute to depression and stress-sensitivity. Gene-based test identified *ZNF366* as the only gene achieving genome-wide significance ($p = 1.48\text{x}10^{-7}$; Bonferroni-corrected significance threshold $p < 2.79\text{x}10^{-6}$; S5 Table and S5 Fig). Using summary statistics

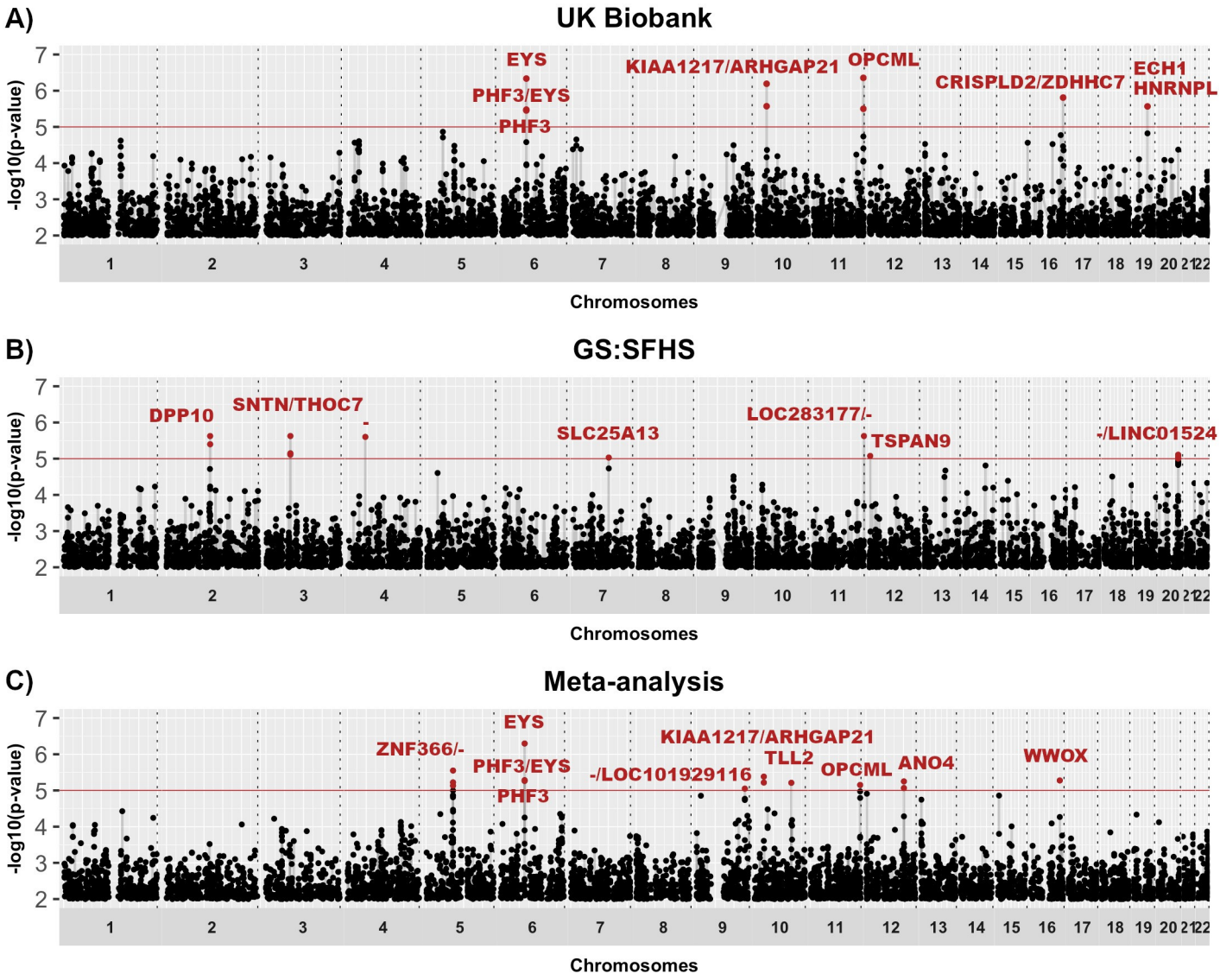


Fig 1. Manhattan plots showing stress-sensitivity associations. Manhattan plots of the GWIS from (A) UKB, (B) GS:SFHS and (C) sample size weighted meta-analysis of UKB and GS:SFHS. The x-axis is chromosomal position and y-axis is the *p* value (-log10 *p* value) of association with stress-sensitivity effect. Suggestive genome-wide significance threshold ($p = 1x10^{-5}$) is shown by solid line at y = 5. Genes or closest gene up- and down-stream from SNP position (/) are annotated. "-": No gene within 100kb of the SNP.

https://doi.org/10.1371/journal.pone.0209160.g001

from meta-analysis GWIS results, stress-sensitivity SNP-based heritability was estimated from LD score regression at 5.0% ($h^2 = 0.0499$, s.e. = 0.017, $p = 1.67x10^{-3}$). Conversely, the SNP-based heritability for MDD and neuroticism were estimated at 9.6% ($h^2 = 0.0962$, s.e. = 0.0179, $p = 3.87x10^{-8}$) and 10.1% ($h^2 = 0.1006$, s.e. = 0.0076, $p = 3.47x10^{-40}$) respectively, using summary statistics from the meta-analysed GWAS of UKB and GS:SFHS.

Pathway enrichment, functional annotation and gene expression analyses

Lead SNPs from the GWIS meta-analysis were investigated using DEPICT. No gene showed statistically significant links to stress-sensitivity at a DEPICT false discovery rate (FDR) < 0.05. No significant result was found for either gene set analysis or tissue enrichment analysis

Table 1. Top 25 SNPs from meta-analysis of GWISs.

Rank	CHR	SNP	BP	A1	Z-score	Effect[a]	p[b]	p (EPQN) [c]	p (MDD) [d]	GENE	POSITION[e]
1	6	rs319924	64487247	A	5.024	++	$5.05x10^{-7}$	0.376	0.637	*EYS*	Intronic
2	5	rs246565	71809247	A	-4.684	—	$2.82x10^{-6}$	0.248	0.589	*ZNF366*	5998bp 5'
3	10	rs2265265	24854876	A	4.604	++	$4.15x10^{-6}$	0.035	0.084	*KIAA1217 / ARHGAP21*	18104bp 3' / 17662bp 3'
4	6	rs1057530	64427095	A	-4.556	—	$5.21x10^{-6}$	0.636	0.840	*PHF3/EYS*	1677bp 3' / 2781bp 3'
5	16	rs7199110	78790765	A	-4.553	—	$5.29x10^{-6}$	0.661	0.741	*WWOX*	Intronic
6	6	rs10485358	64386060	A	-4.546	—	$5.46x10^{-6}$	0.390	0.902	*PHF3*	Intronic
7	12	rs10778077	101193988	A	4.54	++	$5.62x10^{-6}$	0.614	0.430	*ANO4*	Intronic
8	5	rs13358894	71803446	A	4.527	++	$5.99x10^{-6}$	0.257	0.651	*ZNF366*	197bp 5'
9	10	rs2256220	24856314	A	-4.524	—	$6.06x10^{-6}$	0.134	0.129	*KIAA1217 / ARHGAP21*	19542bp 3' / 16224bp 3'
10	10	rs3762096	98136250	A	-4.521	—	$6.15x10^{-6}$	0.437	0.149	*TLL2*	Intronic
11	11	rs2221540	132716369	A	-4.492	—	$7.05x10^{-6}$	0.468	0.364	*OPCML*	Intronic
12	5	rs10043659	71781839	A	4.483	++	$7.37x10^{-6}$	0.339	0.808	*ZNF366*	Intronic
13	12	rs10778078	101195088	A	-4.45	—	$8.58x10^{-6}$	0.599	0.456	*ANO4*	Intronic
14	9	rs10987199	128968987	A	-4.442	—	$8.91x10^{-6}$	0.199	0.026	*LOC101929116*	63416bp 3'
15	5	rs10042132	71789021	A	-4.416	—	$1.01x10^{-5}$	0.418	0.538	*ZNF366*	Intronic
16	11	rs10894606	132671611	A	-4.404	—	$1.06x10^{-5}$	0.438	0.587	*OPCML*	Intronic
17	12	rs7295089	2440464	A	4.372	++	$1.23x10^{-5}$	0.266	0.212	*CACNA1C*	Intronic
18	5	rs9293292	71696942	A	-4.351	—	$1.36x10^{-5}$	0.126	0.731	*PTCD2/ZNF366*	41762bp 3' / 42292bp 3'
19	15	rs3097437	27872136	A	4.346	++	$1.38x10^{-5}$	0.970	0.226	*GABRG3*	93762bp 3'
20	9	rs1999377	11919732	A	4.344	++	$1.40x10^{-5}$	0.436	0.064	-	Intragenic
21	5	rs6862221	71754962	A	4.342	++	$1.41x10^{-5}$	0.543	0.823	*ZNF366*	Intronic
22	5	rs9293289	71683885	A	-4.323	—	$1.54x10^{-5}$	0.395	0.510	*PTCD2/ZNF366*	28705bp 3' / 55349bp 3'
23	11	rs4575282	132719646	A	-4.313	—	$1.61x10^{-5}$	0.598	0.514	*OPCML*	Intronic
24	9	rs2417008	128970219	A	-4.3	—	$1.71x10^{-5}$	0.208	0.026	*LOC101929116*	62184bp 3'
25	9	rs7021461	128972210	A	4.299	++	$1.72x10^{-5}$	0.202	0.025	*LOC101929116*	60193bp 3'

[a]Effect direction in GS:SFHS and UK Biobank.

[b,c,d]Significances of

[b]GWIS stress-sensitivity effect

[c]SNP main effect on neuroticism derived from GWAS meta-analysis of EPQN between UK Biobank and Generation Scotland

[d]SNP main effect on MDD derived from GWAS meta-analysis of MDD between UK Biobank and Generation Scotland.

[e]Position of the SNP respect to closest gene transcripts within 100kb (including UTRs) from 5 prime (5') or 3prime (3').

LD score regression was performed to obtain genetic correlations between stress-sensitivity, MDD and neuroticism. As previously shown, there was a significant genetic correlation between MDD and neuroticism (r_G = 0.637, s.e. = 0.0704, $p = 1.39x10^{-19}$). However, we found no evidence for a genetic correlation between stress-sensitivity and MDD (rG = -0.099, s.e. = 0.182, p = 0.585) or between stress-sensitivity and neuroticism (r_G = 0.114, s.e. = 0.107, p = 0.285).

https://doi.org/10.1371/journal.pone.0209160.t001

at FDR < 0.05. Evidence of regulatory elements on normal cell lines/tissues was identified for 5 of the 12 lead SNPs (i.e. rs3762096, rs10987199, rs2221540, rs246565, rs319924). Two lead SNPs were associated with eQTLs: rs319924 (an intronic SNP in *EYS*) and rs9509508 (an intronic SNP in *LATS2*) and potentially regulate *LGSN/RP3-407E4.3* ($p = 6.31x10^{-12}$/ $p = 1.15x10^{-5}$) and *LATS2* ($p = 3.74x10^{-8}$), respectively.

Polygenic risk scores for stress-sensitivity predict MDD liability

PRS were used to investigate whether common variants affecting stress-sensitivity predict MDD risk. We generated PRS (PRS_{SS}) for stress-sensitivity based on the summary statistics from the GWIS. After 10,000 permutations, PRS_{SS} significantly predicted MDD risk in GS: SFHS using weights from the larger UKB summary data (Empirical-p = 0.04; $p = 5.2x10^{-3}$; β =

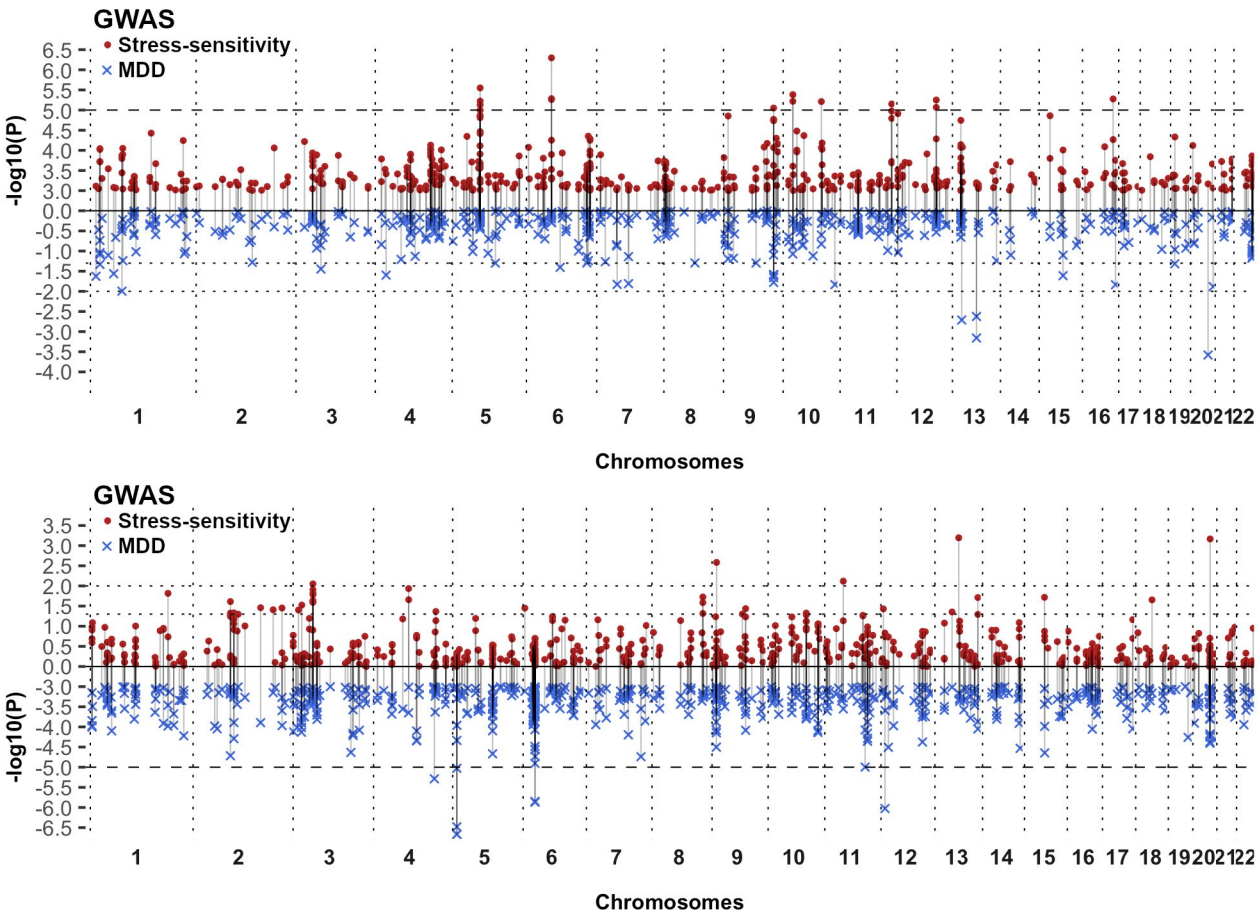


Fig 2. Miami plots showing comparison between association profile between stress-sensitivity GWIS and MDD GWAS. Miami plots from meta-analysis filter at $p = 1x10^{-3}$: (A) filtering for stress-sensitivity p values (•), (B) filtering for MDD p values (×). The x-axis is chromosomal position and y-axis is the p value (-log10 p value) of association with stress-sensitivity (up; red dots) and MDD p value (down; blue crosses). Dot line: genome-wide suggestive threshold ($p = 1x10^{-5}$) at the filtered effect; dashed lines: $p = 0.01$ and 0.05 at unfiltered effect.

https://doi.org/10.1371/journal.pone.0209160.g002

0.078, s.e. = 0.028; best-fit p threshold = 0.005; S6 Table). On the liability scale, the MDD variance explained in GS:SFHS by PRS_{SS} was modest ($R^2 = 0.195\%$). This was less than predicted by PRS weighted by the genetic main effects of MDD or neuroticism (PRS_D: $R^2 = 0.368\%$; PRS_N: $R^2 = 0.459\%$; Table 2 and S6 Table). However, this association was not cross-validated in UKB using summary data from the smaller GS:SFHS GWIS (Empirical-$p = 0.68$; $p = 0.23$; $\beta = 0.004$, s.e. = 0.003; best-fit p threshold = 0.005; PRS_{SS} $R^2 = 0.013\%$; S6 Table), likely due to lack of power as a result of the small discovery sample size. PRS_D ($R^2 = 0.204\%$) and PRS_N ($R^2 = 0.166\%$) derived from GS:SFHS significantly predicted MDD in UKB (Table 2 and S6 Table).

Due to the known genetic correlations between MDD, neuroticism and stressful life events [21], models jointly fitting the effects of multiple PRS were analysed. Multiple regression analyses in GS:SFHS showed that, compared to PRS_D effects alone, the stress-sensitivity effect derived from the UKB GWIS effects significantly explains an additional 0.195% (a predictive improvement of 53.1%, $p = 5.1x10^{-3}$; PRS_D: $\beta = 0.112$, s.e. = 0.029; PRS_{SS}: $\beta = 0.078$, s.e. = 0.028). The inclusion of PRS_{SS} in the full model, where PRS_{SS} was fitted along with both PRS_D and PRS_N weighted by GWAS summary statistics derived from UKB remained significant; explaining an additional 0.172% (a predictive improvement of 24.6%, $p = 8.5x10^{-3}$; PRS_D:

Table 2. MDD risk prediction at best fits.

UKB predicting on GS:SFHS						
Weighted effect	**Best fit threshold**	**# SNPs**	**R^2 (%)[d]**	**R^2 (%)[e]**	***p***	**Empirical-*p***
Stress-sensitivity	0.005	1,626	0.141	0.195	$5.2x10^{-3}$	0.0399
MDD[a]	0.1	22,771	0.265	0.368	$1.3x10^{-4}$	0.0015
EPQN[b]	0.4	65,276	0.330	0.459	$1.8x10^{-5}$	0.0002
MDD[a] + EPQN[b]	-	-	0.503	0.699	$8.0x10^{-7}$	-
joint models[c]	-	-	0.627	0.871	$1.2x10^{-7}$	-
PGC2 & GPC predicting on GS:SFHS						
PGC2 MDD[a]	1	92,248	0.993	1.378	$1.4x10^{-13}$	≤0.0001
GPC EPQN[b]	0.01	3,521	0.108	0.149	0.014	0.1038
PGC2 MDD + GPC EPQN[b]	-	-	1.052	1.461	$1.7x10^{-13}$	-
joint models[c]	-	-	1.203	1.671	$1.6x10^{-14}$	-
GS:SFHS predicting on UKB						
Weighted effect	**Best fit threshold**	**# SNPs**	**R^2 (%)[a]**	**R^2 (%)[b]**	***p***	**Empirical-*p***
Stress-sensitivity	0.005	1,526	0.008	0.013	0.231	0.6841
MDD[a]	0.03	7,725	0.130	0.204	$1.6x10^{-6}$	≤0.0001
EPQN[b]	0.05	12,296	0.106	0.166	$1.6x10^{-5}$	0.0005
MDD[a] + EPQN[b]	-	-	0.197	0.309	$2.8x10^{-8}$	-
joint models[c]	-	-	0.206	0.322	$6.6x10^{-8}$	-
PGC2 & GPC predicting on UKB						
PGC2 MDD[a]	0.5	64,113	0.919	1.440	$3.4x10^{-37}$	<0.0001
GPC EPQN[b]	0.03	8,761	0.066	0.104	$6.5x10^{-4}$	0.006
PGC2 MDD[a] + GPC EPQN[b]	-	-	0.950	1.488	$2.9x10^{-37}$	-
joint models[c]	-	-	0.958	1.501	$1.5x10^{-36}$	-

[a]major depressive disorder

[b]neuroticism score

[c]combined effect fitting all 3 PRS weighted by all the effects (i.e. stress-sensitivity, MDD and EPQN)

[d]Nagelkerke's R^2 at observed scale

[e]R^2 on the liability scale.

https://doi.org/10.1371/journal.pone.0209160.t002

β = 0.093, s.e. = 0.029; PRS_N: β = 0.107, s.e. = 0.030; PRS_{SS}: β = 0.073, s.e. = 0.028). In models fitting PRS_D and PRS_N, the variances explained were non-additive, demonstrating the partial overlap between MDD risk prediction from PRS_D and PRS_N main additive effects. This is consistent with the known genetic correlation between these two traits. An overlap was not seen between the variance explained by PRS_{SS} effect and the variance explained by PRS_D and/or PRS_N. Multiple regression analyses fitting PRS_D and PRS_N derived from worldwide consortiums (Fig 3) showed that the increased sample size from GWAS used to derive PRS_D resulted in an increment of MDD variance explained in GS:SFHS by PRS_D (from 0.368% to 1.378%). However, there was no change in the proportion of the variance explained by the PRS_{SS} in the full model (PRS_{SS} $p = 3.5x10^{-3}$). These results suggest that PRS_{SS} explains a proportion of MDD risk not accounted for by PRS_D or PRS_N at current sample sizes. However, these findings were not cross-validated in UKB using PRS_{SS} derived from GS:SFHS GWIS, likely due to lack of power as a result of the small discovery sample size (S6 Fig).

Using stress-sensitivity to stratify MDD in GS:SFHS

MDD cases show significantly higher PRS_{SS} ($p = 2x10^{-3}$) and PRS_D ($p = 1.8x10^{-4}$) than controls. Association between MDD-related traits and stress-sensitivity risk quintiles was assessed

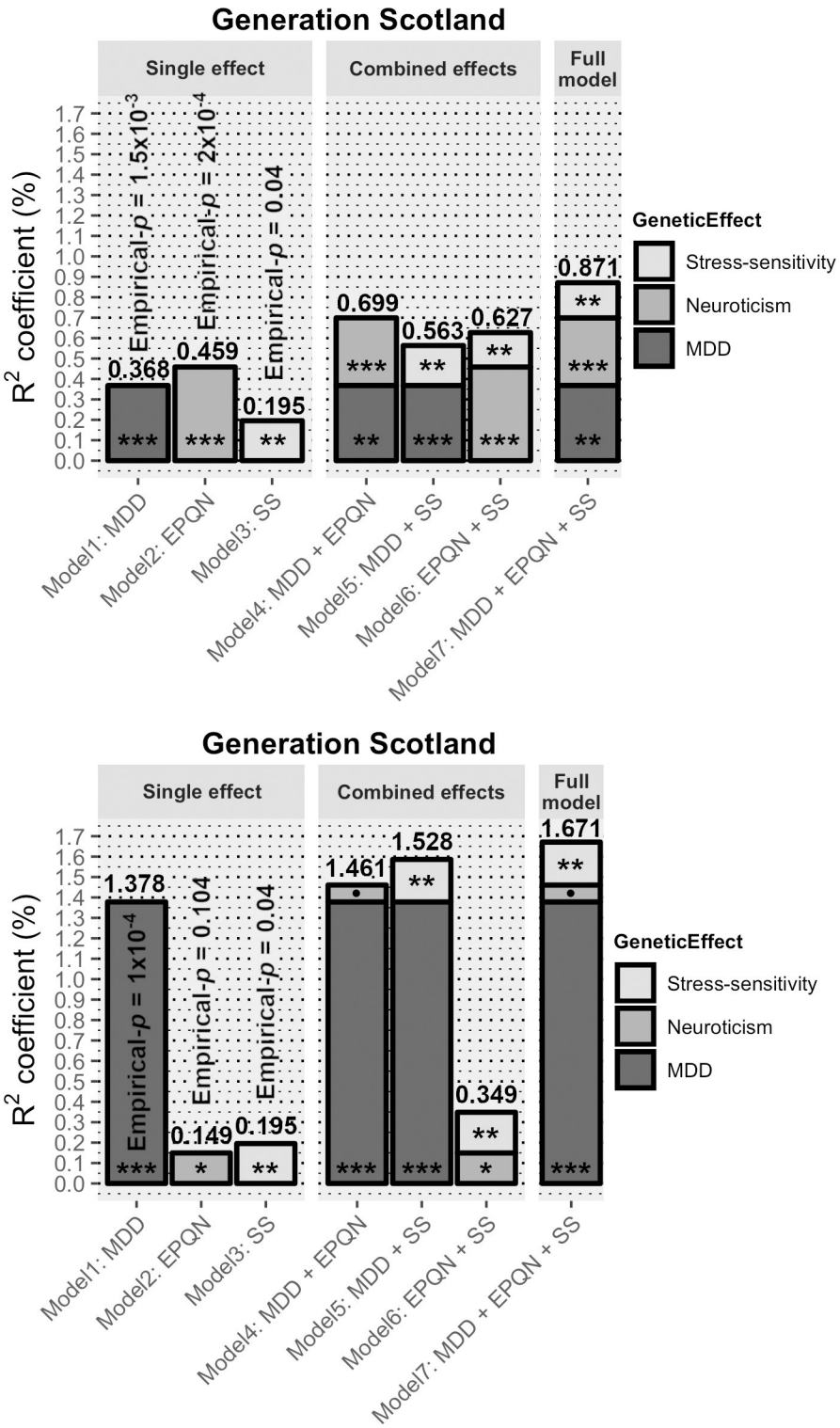


Fig 3. MDD is best predicted using multiple PRS. MDD risk explained (R^2 coefficient (%); top bar values) on the liability scale by each PRS in GS:SFHS; weighted by GWAS main additive and GWIS stress-sensitivity effects independently and combined. (A) Using summary statistics from UKB as discovery sample. There is an increment on MDD risk prediction from adding PRS_{SS} to PRS_D model of 53.1% and 24.6% when combining PRS_{SS} with both MDD and neuroticism PRS. (B) Replication of fitting PRS_D and PRS_N using summary statistics from worldwide consortiums (i.e. PGC &GPC). Significance codes: *p* values $^{***} < 0.001 < ^{**} < 0.01 < ^{*} < 0.05 < \bullet < 0.1$; derived from likelihood ratio tests. *SS* stands for stress-sensitivity.

https://doi.org/10.1371/journal.pone.0209160.g003

on MDD cases in order to identify a subgroup of MDD patients, perhaps defining a characteristic aetiological subtype of MDD. However, stratification analysis failed, and no quintile based on PRS_{SS} nor its interaction with quintiles based on PRS_D showed statistically significant effects on any trait analyzed. Individuals with high PRS_{SS} were not significantly different from other cases for sex, MDD course, age of onset or episode count, nor neuroticism, mood disorder or schizotypal personality scores ($p > 0.05$; S7 Table). Results remained non significant when PRSs were fitted as continuous variables ($p > 0.05$).

Discussion

The existence of genetic variants affecting an individual's risk of depression in response to stress has been predicted previously [46, 49, 50] and is consistent with the departure from a simple additive genetic model seen in twin-studies of recurrent depressive disorder [83]. Through international research efforts such as the PGC and UK Biobank, there are ever-increasing sample sizes available for understanding the genetics of MDD. These resources are beginning, and will continue to, identify genome-wide significant loci [42, 84, 85]. However, the lack of environmental data and/or their reliability, makes the study of genetic individual's response to their negative effects, and their contribution to the onset of MDD and other stress-related disorders, difficult. As a way to address this limitation, we generated a proxy for stress-sensitivity through modelling the interaction between SNP allele and MDD status on neuroticism score in a GWIS approach. Thus, we sought to identify the genetic underpinnings of individual's sensitivity to stress response (stress-sensitivity) through those variants that contribute to higher neuroticism levels only in individuals with a lifetime diagnosis of MDD but not in healthy controls.

We performed a GWIS to identify loci showing differential effects on neuroticism scores in individuals with and without MDD (so called stress-sensitivity proxy). No SNPs reached genome-wide significance, but 14 SNPs from 8 loci reached suggestive significance levels (see S4 Table for prior evidence of associated phenotypes). Enrichment analysis showed no evidence for enrichment of specific pathways or tissues. The top two loci, *PTP4A1-PHF3-EYS* and *ZNF366* have been previously associated with alcohol dependence [86–90], alcohol intake (dbGaP: phs000342) and glucocorticoid receptor function [91–93]. The most significant SNP in this study, rs319924, is an intronic variant in *EYS* that is a potential eQTL for *LGSN* [76], a gene previously associated with male-specific depression [94]. This is of particular interest given previous studies linking alcohol consumption, stress and the risk of depression [95–100]. However, findings should be interpreted with caution, as these loci did not reach genome-wide significance at current sample size. Evidence of an eQTL effect was predicted for a lead SNP in *LATS2*, a positive regulator of histone methyltransferase activity [101] a process important in anxiety-related behaviours [102]. The prior association of the top two loci in this study with alcohol related-phenotypes suggests that genes involved in the sensitivity to stress may mediate the effects of stress on alcohol consumption. Some *PHF3* paralogs have been shown to be linked with depression and modulate stress response [103, 104].

Gene-based analysis identified a genome-wide significant association between *ZNF366* and stress-sensitivity. *ZNF366* (also known as *DC-SCRIPT*) is a corepressor of transcription found in nuclear receptor complexes including the glucocorticoid receptor. *ZNF366* represses glucocorticoid receptor-mediated transcription in monocyte-derived dendritic cells [91]; and may act through histone deacetylases to modulate immune response [92]. There is evidence from a large-scale mRNA display study that *PHF3*, in the region underlying the most significant peak in the single SNP analysis, may also interact, directly or indirectly, with the glucocorticoid receptor (IntAct database [93]) but this has not been confirmed. These results reinforce the

hypothesis that our proxy for stress-sensitivity truly reflects the genetic architecture of sensitivity to respond to stress.

We estimated a significant lower bound on common SNP-based heritability for stress-sensitivity of 5%. Whilst the known genetic overlap between MDD and neuroticism was detectable, the lack of genetic correlation with stress-sensitivity, reinforced by results from multiple regression analyses, indicated a lack of significant overlap in the genetics factors underpinning stress-sensitivity and MDD or neuroticism. This analysis may be limited by our sample size, although using the largest available meta-analyses of MDD and neuroticism [42, 77] did not decrease the proportion of liability explained by the PRS_{SS}. We note, that as such meta-analyses increase in size it is likely, as with the effects of smoking in schizophrenia [105, 106], that the indirect genetic effects of the environment on the risk of depression will be detected by GWAS. However, through studies such as ours, or similar, the mechanism for the effect of the risk alleles may be clarified.

Further, we show that such genetic information in stress-sensitivity could significantly improve the proportion of liability to MDD predicted by PRS based only on additive genetic effects on MDD identified by large GWAS. The summary results from the GWIS were used to derive a PRS reflecting the genetic difference in stress-sensitivity. This variable significantly predicted liability to MDD in GS:SFHS ($p = 5.2\text{x}10^{-3}$, Empirical-$p = 0.04$ after 10,000 permutations), although this finding could not be replicated in UKB (Empirical-$p = 0.68$), likely due to lack of power. This is consistent with the expectation that the larger the discovery sample (i.e. UKB), the greater the accuracy of the weighting and the more predictive the PRS [107]. Multiple regression models in GS:SFHS suggest that inclusion of PRS weighted by stress-sensitivity significantly improves MDD prediction over use of either MDD and/or neuroticism weighted PRS alone (improvement in full model $p = 8.5\text{x}10^{-3}$). However, we were unable to identify a subgroup of MDD cases with higher PRS_{ss}. The polygenic interaction approach used in our study may, therefore, improve the interpretation of both positive and negative findings from GWAS studies (i.e. pathways and mechanisms involved, lack of replication, or negative findings in variants mediating environmental effects). Added to paralleling recent developments in GWAS analyses, it may maximize our power to detect gene-by-environment effects in this heterogeneous disorder.

Future studies will be required to further investigate the effects of adverse life events in individuals with high or low polygenic risk scores for stress-sensitivity. However, the methodology presented allows addressing the genetic response to negative outcomes via proxy in the absence of prospective environmental data.

Here we identify an independent set of risk variants for an individual's response to negative outcomes and show that incorporating information across many loci provides clear and replicable evidence for a genetic effect of stress-sensitivity on MDD risk; identifying a potential genetic link with alcohol intake. These results require further study, but may inform treatment of comorbid alcohol dependency and depression.

Supporting information

S1 Supporting Information.
(DOCX)

S1 File. GWIS summary statistics from Generation Scotland.
(CSV)

S2 File. GWIS summary statistics from UK Biobank.
(CSV)

S3 File. GWIS summary statistics from meta-analysis.
(CSV)

S1 Fig. Genetic stress-sensitivity effect representation. Genetic stress-sensitivity effect on MDD (β_{SS}) is defined as the difference between the regression coefficient in MDD cases (β_A) and the regression coefficient in controls (β_B) from linear models regressed on EPQN, adjusted by covariates. *A1*: allele 1. *A2* allele 2.
(TIFF)

S2 Fig. QQ plot from stress-sensitivity meta-analysis. QQ plot of GWIS from sample size weighted meta-analysis ($\lambda = 0.997$; s.e. = 1.05×10^{-5}). All SNPs wit $p < 2 \times 10^{-5}$, p threshold (dot line) where some SNPs start to deviate from null distribution going outside 95% confidence intervals (grey shadow), were selected to perform DEPICT analyses to assess pathway and functional genomic analyses. 27 top variants from 12 independent loci were selected.
(TIFF)

S3 Fig. QQ plots of GWIS *p* values. QQ plots of GWIS from (A) UKB ($\lambda = 1.014$; s.e. = 1.027×10^{-5}), (B) GS:SFHS ($\lambda = 0.997$; s.e. = 7.989×10^{-6}). The 95% confidence interval is shaded in grey.
(TIFF)

S4 Fig. Miami plots on UK Biobank and Generation Scotland: Scottish Family Health Study. Miami plots showing comparison between association profile between SS and MDD main additive effects. Miami plots from (A) UKB filtering for SS p values (top) and MDD p values (bottom), (B) GS:SFHS filtering for SS p values (top) and MDD p values (bottom). Filter at $p = 1 \times 10^{-3}$. The x-axis is base-paired chromosomal position and y-axis is the significance (-log10 p) of association with (up; red dots) SS effect and (down; blue dots) MDD. Dot line: genome-wide suggestive threshold ($p = 1 \times 10^{-5}$) at the filtered effect; dashes lines: p value = 0.01 and 0.05 at compared effect.
(TIFF)

S5 Fig. Manhattan plot of the gene-based test for stress-sensitivity. Manhattan plot showing gene-based association of stress-sensitivity. The x-axis is base-paired chromosomal position and y-axis is the significance ($-\log_{10} p$ value) of association with SS effect. Genome-wide significance threshold showed by red dashed line was defined at $p = 0.05/17{,}931 = 2.79 \times 10^{-6}$.
(TIFF)

S6 Fig. PRS profiling predicting MDD in UK Biobank. MDD risk explained (R^2 coefficient (%); top bar values) on the liability scale by each PRS in UKB; weighted by GWAS main additive and GWIS stress-sensitivity effects independently and combined. (A) Using summary statistics from GS:SFHS as discovery sample. (B) Replication fitting PRS_D and PRS_N using summary statistics from worldwide consortiums (i.e. PGC & GPC). Significance codes: p values $^{***} < 0.001 < ^{**} < 0.01 < ^{*} < 0.05$; derived from likelihood ratio tests. *SS* stands for stress-sensitivity.
(TIFF)

S1 Table. EPQN comparison between MDD cases and healthy controls.
(XLSX)

S2 Table. Top 10 SNPs from GWIS on UK Biobank.
(XLSX)

S3 Table. Top 10 SNPs from GWIS on Generation Scotland: Scottish Family Health Study.
(XLSX)

S4 Table. Traits with significant evidence of association with closest gene to suggestive stress-sensitive hits. The closest genes to SNPs associated with stress-sensitivity at suggestive significance levels have prior evidence of association in dbGAP with a wide range of neuropsychiatric traits such as schizophrenia, bipolar disorder, attention deficit disorder with hyperactivity, mental competency, intuition, sleep or alcohol drinking.
(XLSX)

S5 Table. Top 25 hits from gene-based analysis of GWIS meta-analysis.
(XLSX)

S6 Table. Summary results from polygenic risk score (PRS) analysis using PRSice-2.
(XLSX)

S7 Table. MDD stratification.
(XLSX)

Acknowledgments

We thank all participants from UK Biobank, including families, volunteers, technicians and scientist who helped collect the samples and prepare the data. Analyses were performed under UK Biobank project 4844. We are grateful to all the families who took part in Generation Scotland, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, the **Generation Scotland** members (David Porteous, Archie Campbell, Blair H Smith, Corri Black, Sandosh Padmanabhan, Caroline Hayward, Andrew McIntosh and Ian J Deary), the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. We acknowledge the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh to carry out the genotyping of the Generation Scotland samples. We gratefully acknowledge **The Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**, which depends on the contributions of many parties: Naomi R Wray (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU and Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU), Stephan Ripke (Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US, Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, DE and Medical and Population Genetics, Broad Institute, Cambridge, MA, US), Manuel Mattheisen (Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, SE, Department of Biomedicine, Aarhus University, Aarhus, DK, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK and iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK), Maciej Trzaskowski (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD), Enda M Byrne (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU), Abdel Abdellaoui (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Mark J Adams (Division of Psychiatry, University of Edinburgh, Edinburgh, GB), Esben Agerbo (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK and National Centre for Register-Based Research, Aarhus University, Aarhus, DK), Tracy M Air (Discipline of Psychiatry, University of Adelaide, Adelaide, SA, AU), Till F M Andlauer

(Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE and Munich Cluster for Systems Neurology (SyNergy), Munich, DE), Silviu-Alin Bacanu (Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US), Marie Bækvad-Hansen (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK), Aartjan T F Beekman (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Tim B Bigdeli (Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US and Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, US), Elisabeth B Binder (Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE and Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, US), Douglas H R Blackwood (Division of Psychiatry, University of Edinburgh, Edinburgh, GB), Julien Bryois (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE), Henriette N Buttenschøn (iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK, iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, DK), Jonas Bybjerg-Grauholm (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK), Na Cai (Human Genetics, Wellcome Trust Sanger Institute, Cambridge, GB and Statistical genomics and systems genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, GB), Enrique Castelao (Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH), Jane Hvarregaard Christensen (Department of Biomedicine, Aarhus University, Aarhus, DK, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK and iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK), Toni-Kim Clarke (Division of Psychiatry, University of Edinburgh, Edinburgh, GB), Jonathan R I Coleman (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Lucía Colodro-Conde (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU), Baptiste Couvy-Duchesne (Centre for Advanced Imaging, The University of Queensland, Saint Lucia, QLD, AU and Queensland Brain Institute, The University of Queensland, Saint Lucia, QLD, AU), Nick Craddock (Psychological Medicine, Cardiff University, Cardiff, GB), Gregory E Crawford (Center for Genomic and Computational Biology, Duke University, Durham, NC, US and Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, US), Gail Davies (Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB), Ian J Deary (Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB), Franziska Degenhardt (Institute of Human Genetics, University of Bonn, Bonn, DE and Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE), Eske M Derks (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU), Nese Direk (Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL and Psychiatry, Dokuz Eylul University School Of Medicine, Izmir, TR), Conor V Dolan (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Erin C Dunn (Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US, Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, US and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, US), Thalia C Eley (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Valentina Escott-Price (Neuroscience and Mental Health, Cardiff University, Cardiff, GB), Farnush Farhadi Hassan Kiadeh (Bioinformatics, University of British Columbia,

Vancouver, BC, CA), Hilary K Finucane (Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, US and Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, US), Andreas J Forstner (Institute of Human Genetics, University of Bonn, Bonn, DE, Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE, Department of Psychiatry (UPK), University of Basel, Basel, CH and Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH), Josef Frank (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Héléna A Gaspar (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Michael Gill (Department of Psychiatry, Trinity College Dublin, Dublin, IE), Fernando S Goes (Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US), Scott D Gordon (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU), Jakob Grove (Department of Biomedicine, Aarhus University, Aarhus, DK, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK, iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Bioinformatics Research Centre, Aarhus University, Aarhus, DK), Lynsey S Hall (Division of Psychiatry, University of Edinburgh, Edinburgh, GB and Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, GB), Christine Søholm Hansen (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK), Thomas F Hansen (Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, DK, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK and iPSYCH, The Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, DK), Stefan Herms (Institute of Human Genetics, University of Bonn, Bonn, DE, Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE and Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH), Ian B Hickie (Brain and Mind Centre, University of Sydney, Sydney, NSW, AU), Per Hoffmann (Institute of Human Genetics, University of Bonn, Bonn, DE, Life&Brain Center, Department of Genomics, University of Bonn, Bonn, DE and Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH), Georg Homuth (Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Carsten Horn (Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH), Jouke-Jan Hottenga (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), David M Hougaard (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK), Marcus Ising (Max Planck Institute of Psychiatry, Munich, DE), Rick Jansen (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Eric Jorgenson (Division of Research, Kaiser Permanente Northern California, Oakland, CA, US), James A Knowles (Psychiatry & The Behavioral Sciences, University of Southern California, Los Angeles, CA, US), Isaac S Kohane (Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US, Department of Medicine, Brigham and Women's Hospital, Boston, MA, US and Informatics Program, Boston Children's Hospital, Boston, MA, US), Julia Kraft (Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, DE), Warren W. Kretzschmar (Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, GB), Jesper Krogh (Department of Endocrinology at Herlev University

Hospital, University of Copenhagen, Copenhagen, DK), Zoltán Kutalik (Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, VD, CH and Swiss Institute of Bioinformatics, Lausanne, VD, CH), Yihan Li (Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, GB), Penelope A Lind (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU), Donald J MacIntyre Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, GB and Mental Health, NHS 24, Glasgow, GB), Dean F MacKinnon (Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US), Robert M Maier (Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU), Wolfgang Maier (Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, DE), Jonathan Marchini (Statistics, University of Oxford, Oxford, GB), Hamdi Mbarek (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Patrick McGrath (Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, US), Peter McGuffin (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Sarah E Medland (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU), Divya Mehta (Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU and School of Psychology and Counseling, Queensland University of Technology, Brisbane, QLD, AU), Christel M Middeldorp (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL, Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, QLD, AU and Child Health Research Centre, University of Queensland, Brisbane, QLD, AU), Evelin Mihailov (Estonian Genome Center, University of Tartu, Tartu, EE), Yuri Milaneschi (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Lili Milani (Estonian Genome Center, University of Tartu, Tartu, EE), Francis M Mondimore (Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US), Grant W Montgomery (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU), Sara Mostafavi (Medical Genetics, University of British Columbia, Vancouver, BC, CA and Statistics, University of British Columbia, Vancouver, BC, CA), Niamh Mullins (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Matthias Nauck DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE and Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Bernard Ng (Statistics, University of British Columbia, Vancouver, BC, CA), Michel G Nivard (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Dale R Nyholt (Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, AU), Paul F O'Reilly (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB), Hogni Oskarsson (Humus, Reykjavik, IS), Michael J Owen (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB), Jodie N Painter (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU), Carsten Bøcker Pedersen (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK and National Centre for Register-Based Research, Aarhus University, Aarhus, DK), Marianne Giørtz Pedersen (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK and National Centre for Register-Based Research, Aarhus University, Aarhus, DK), Roseann E. Peterson (Department of Psychiatry, Virginia Commonwealth University,

Richmond, VA, US and Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US), Erik Pettersson (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE), Wouter J Peyrot (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Giorgio Pistis (Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH), Danielle Posthuma (Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, NL and Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, NL), Jorge A Quiroz (Solid Biosciences, Boston, MA, US), Per Qvist (Department of Biomedicine, Aarhus University, Aarhus, DK, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK and iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK), John P Rice (Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, US), Brien P. Riley (Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US), Margarita Rivera (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB and Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, ES), Saira Saeed Mirza (Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL), Robert Schoevers (Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, NL), Eva C Schulte (Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, DE and Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, DE), Ling Shen (Division of Research, Kaiser Permanente Northern California, Oakland, CA, US), Jianxin Shi (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US), Stanley I Shyn (Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, US), Engilbert Sigurdsson (Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, IS), Grant C B Sinnamon (School of Medicine and Dentistry, James Cook University, Townsville, QLD, AU), Johannes H Smit (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Daniel J Smith (Institute of Health and Wellbeing, University of Glasgow, Glasgow, GB), Hreinn Stefansson (deCODE Genetics / Amgen, Reykjavik, IS), Stacy Steinberg (deCODE Genetics / Amgen, Reykjavik, IS), Fabian Streit (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Jana Strohmaier (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Katherine E Tansey (College of Biomedical and Life Sciences, Cardiff University, Cardiff, GB), Henning Teismann (Institute of Epidemiology and Social Medicine, University of Münster, Münster, Nordrhein-Westfalen, DE), Alexander Teumer (Institute for Community Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Wesley Thompson iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK, Department of Psychiatry, University of California, San Diego, San Diego, CA, US and KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, NO), Pippa A Thomson (Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, GB), Thorgeir E Thorgeirsson (deCODE Genetics / Amgen, Reykjavik, IS), Matthew Traylor (Clinical Neurosciences, University of Cambridge, Cambridge, GB), Jens Treutlein (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Vassily Trubetskoy (Department of Psychiatry and Psychotherapy, Universitätsmedizin

Berlin Campus Charité Mitte, Berlin, DE), André G Uitterlinden (Internal Medicine, Erasmus MC, Rotterdam, Zuid-Holland, NL), Daniel Umbricht (Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery & Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH), Sandra Van der Auwera (Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Albert M van Hemert (Department of Psychiatry, Leiden University Medical Center, Leiden, NL), Alexander Viktorin (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE), Peter M Visscher (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU and Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU), Yunpeng Wang (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK and KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, NO), Bradley T. Webb (Virginia Institute of Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US), Shantel Marie Weinsheimer (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK), Jürgen Wellmann (Institute of Epidemiology and Social Medicine, University of Münster, Münster, Nordrhein-Westfalen, DE), Gonneke Willemsen (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Stephanie H Witt (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Yang Wu (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU), Hualin S Xi (Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, US), Jian Yang (Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU and Institute for Molecular Bioscience; Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU), Futao Zhang (Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU), Volker Arolt (Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE), Bernhard T Baune (Discipline of Psychiatry, University of Adelaide, Adelaide, SA, AU), Klaus Berger (Institute of Epidemiology and Social Medicine, University of Münster, Münster, Nordrhein-Westfalen, DE), Dorret I Boomsma (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL), Sven Cichon (Institute of Human Genetics, University of Bonn, Bonn, DE, Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH, Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, CH and Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, DE), Udo Dannlowski (Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE), EJC de Geus (Dept of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL and Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, NL), J Raymond DePaulo (Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US), Enrico Domenici (Centre for Integrative Biology, Università degli Studi di Trento, Trento, Trentino-Alto Adige, IT), Katharina Domschke (Department of Psychiatry and Psychotherapy, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, DE), Tõnu Esko 5, (Estonian Genome Center, University of Tartu, Tartu, EE), Hans J Grabe (Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Steven P Hamilton (Psychiatry, Kaiser

Permanente Northern California, San Francisco, CA, US), Caroline Hayward (Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, GB), Andrew C Heath (Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, US), Kenneth S Kendler (Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US), Stefan Kloiber (Max Planck Institute of Psychiatry, Munich, DE, Department of Psychiatry, University of Toronto, Toronto, ON, CA and Centre for Addiction and Mental Health, Toronto, ON, CA), Glyn Lewis (Division of Psychiatry, University College London, London, GB), Qingqin S Li (Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, US), Susanne Lucae (Max Planck Institute of Psychiatry, Munich, DE), Pamela AF Madden (Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, US), Patrik K Magnusson (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE), Nicholas G Martin (Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU), Andrew M McIntosh* (Division of Psychiatry, University of Edinburgh, Edinburgh, GB and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB), Andres Metspalu (Estonian Genome Center, University of Tartu, Tartu, EE and Institute of Molecular and Cell Biology, University of Tartu, Tartu, EE), Ole Mors (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, DK), Preben Bo Mortensen (iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK, iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK and National Centre for Register-Based Research, Aarhus University, Aarhus, DK), Bertram Müller-Myhsok (Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE, Munich Cluster for Systems Neurology (SyNergy), Munich, DE and University of Liverpool, Liverpool, GB), Merete Nordentoft (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK and Mental Health Center Copenhagen, Copenhagen Universtity Hospital, Copenhagen, DK), Markus M Nöthen (Institute of Human Genetics, University of Bonn, Bonn, DE and Life&-Brain Center, Department of Genomics, University of Bonn, Bonn, DE), Michael C O'Donovan (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB), Sara A Paciga (Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, US), Nancy L Pedersen (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE), Brenda WJH Penninx (Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL), Roy H Perlis (Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US and Psychiatry, Harvard Medical School, Boston, MA, US), David J Porteous (Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, GB), James B Potash (Psychiatry, University of Iowa, Iowa City, IA, US), Martin Preisig (Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH), Marcella Rietschel (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE), Catherine Schaefer (Division of Research, Kaiser Permanente Northern California, Oakland, CA, US), Thomas G Schulze (Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE, Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, DE, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US, Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Goettingen, Niedersachsen, DE and Human Genetics Branch,

NIMH Division of Intramural Research Programs, Bethesda, MD, US), Jordan W Smoller (Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US, Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, US and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, US), Kari Stefansson (deCODE Genetics / Amgen, Reykjavik, IS and Faculty of Medicine, University of Iceland, Reykjavik, IS), Henning Tiemeier (Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL, Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL and Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL), Rudolf Uher (Psychiatry, Dalhousie University, Halifax, NS, CA), Henry Völzke (Institute for Community Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE), Myrna M Weissman (Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, US and Division of Epidemiology, New York State Psychiatric Institute, New York, NY, US), Thomas Werge (iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK and Department of Clinical Medicine, University of Copenhagen, Copenhagen, DK), Cathryn M Lewis* (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB and Department of Medical & Molecular Genetics, King's College London, London, GB), Douglas F Levinson (Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, US), Gerome Breen (MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB and NIHR BRC for Mental Health, King's College London, London, GB), Anders D Børglum (Department of Biomedicine, Aarhus University, Aarhus, DK, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK and iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, DK), Patrick F Sullivan (Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE, Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US and Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, US). **PGC-MDD group Chairs*:** Catherine Lewis (cathryn.lewis@kcl.ac.uk) and Andrew McIntosh (andrew.mcintosh@ed.ac.uk).

Author Contributions

Conceptualization: Aleix Arnau-Soler, Pippa A. Thomson.

Data curation: Aleix Arnau-Soler, Mark J. Adams.

Formal analysis: Aleix Arnau-Soler, Pippa A. Thomson.

Investigation: Aleix Arnau-Soler, Pippa A. Thomson.

Supervision: Caroline Hayward, Pippa A. Thomson.

Visualization: Aleix Arnau-Soler.

Writing – original draft: Aleix Arnau-Soler.

Writing – review & editing: Mark J. Adams, Caroline Hayward, Pippa A. Thomson.

References

1. Monroe SM. Modern approaches to conceptualizing and measuring human life stress. Annu Rev Clin Psychol. 2008; 4:33–52. https://doi.org/10.1146/annurev.clinpsy.4.022007.141207 PMID: 17716038.
2. Jeronimus BF, Riese H, Sanderman R, Ormel J. Mutual reinforcement between neuroticism and life experiences: a five-wave, 16-year study to test reciprocal causation. Journal of personality and social psychology. 2014; 107(4):751–64. https://doi.org/10.1037/a0037009 PMID: 25111305.

3. Riese H, Snieder H, Jeronimus BF, Korhonen T, Rose RJ, Kaprio J, et al. Timing of Stressful Life Events Affects Stability and Change of Neuroticism. European Journal of Personality. 2014; 28 (2):193–200. https://doi.org/10.1002/per.1929
4. Jeronimus BF, Ormel J, Aleman A, Penninx BW, Riese H. Negative and positive life events are associated with small but lasting change in neuroticism. Psychol Med. 2013; 43(11):2403–15. https://doi.org/10.1017/S0033291713000159 PMID: 23410535.
5. Kendler KS, Kuhn J, Prescott CA. The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. Am J Psychiatry. 2004; 161(4):631–6. https://doi.org/10.1176/appi.ajp.161.4.631 PMID: 15056508.
6. Kessler RC. The effects of stressful life events on depression. Annu Rev Psychol. 1997; 48:191–214. https://doi.org/10.1146/annurev.psych.48.1.191 PMID: 9046559.
7. Tennant C. Life events, stress and depression: a review of recent findings. Aust N Z J Psychiatry. 2002; 36(2):173–82. https://doi.org/10.1046/j.1440-1614.2002.01007.x PMID: 11982537.
8. Porcelli B, Pozza A, Bizzaro N, Fagiolini A, Costantini MC, Terzuoli L, et al. Association between stressful life events and autoimmune diseases: A systematic review and meta-analysis of retrospective case-control studies. Autoimmun Rev. 2016; 15(4):325–34. https://doi.org/10.1016/j.autrev.2015.12.005 PMID: 26708168.
9. Lin Y, Wang C, Zhong Y, Huang X, Peng L, Shan G, et al. Striking life events associated with primary breast cancer susceptibility in women: a meta-analysis study. J Exp Clin Cancer Res. 2013; 32(1):53. https://doi.org/10.1186/1756-9966-32-53 PMID: 23941600; PubMed Central PMCID: PMCPMC3751759.
10. Jafri SHR, Ali F, Mollaeian A, Hasan SM, Hussain R, Akkanti BH, et al. Major stressful life events and risk of developing lung cancer. Journal of Clinical Oncology. 2017; 35(15_suppl):1575–. https://doi.org/10.1200/JCO.2017.35.15_suppl.1575
11. Hasler G, Drevets WC, Manji HK, Charney DS. Discovering endophenotypes for major depression. Neuropsychopharmacology. 2004; 29(10):1765–81. https://doi.org/10.1038/sj.npp.1300506 PMID: 15213704.
12. Bogdan R, Nikolova YS, Pizzagalli DA. Neurogenetics of depression: a focus on reward processing and stress sensitivity. Neurobiol Dis. 2013; 52:12–23. https://doi.org/10.1016/j.nbd.2012.05.007 PMID: 22659304; PubMed Central PMCID: PMCPMC3570616.
13. Lazarus RS, Folkman S. Transactional theory and research on emotions and coping. European Journal of Personality. 1987; 1(3):141–69. https://doi.org/10.1002/per.2410010304
14. Smith MA, Riccalton VC, Kelly-Hughes DH, Craw OA, Allen SF, O'Connor DB, et al. The relationship between Type D personality and physical health complaints is mediated by perceived stress and anxiety but not diurnal cortisol secretion. Stress. 2018:1–8. https://doi.org/10.1080/10253890.2018.1435637 PMID: 29402161.
15. Herr RM, Barrech A, Riedel N, Gundel H, Angerer P, Li J. Long-Term Effectiveness of Stress Management at Work: Effects of the Changes in Perceived Stress Reactivity on Mental Health and Sleep Problems Seven Years Later. Int J Environ Res Public Health. 2018; 15(2). https://doi.org/10.3390/ijerph15020255 PMID: 29401657.
16. Moore RC, Eyler LT, Mausbach BT, Zlatar ZZ, Thompson WK, Peavy G, et al. Complex interplay between health and successful aging: role of perceived stress, resilience, and social support. Am J Geriatr Psychiatry. 2015; 23(6):622–32. https://doi.org/10.1016/j.jagp.2014.08.004 PMID: 25217186; PubMed Central PMCID: PMCPMC4329284.
17. Hayman LW Jr., Lucas T, Porcerelli JH. Cognitive appraisal vs. exposure-based stress measures: links to perceived mental and physical health in low-income black women. J Nerv Ment Dis. 2014; 202 (11):807–12. https://doi.org/10.1097/NMD.0000000000000198 PMID: 25275345.
18. Harris ML, Loxton D, Sibbritt DW, Byles JE. The influence of perceived stress on the onset of arthritis in women: findings from the Australian Longitudinal Study on women's health. Ann Behav Med. 2013; 46(1):9–18. https://doi.org/10.1007/s12160-013-9478-6 PMID: 23436274.
19. Caspi A, Hariri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. Am J Psychiatry. 2010; 167(5):509–27. https://doi.org/10.1176/appi.ajp.2010.09101452 PMID: 20231323; PubMed Central PMCID: PMCPMC2943341.
20. Bleys D, Luyten P, Soenens B, Claes S. Gene-environment interactions between stress and 5-HTTLPR in depression: A meta-analytic update. J Affect Disord. 2018; 226:339–45. https://doi.org/10.1016/j.jad.2017.09.050 PMID: 29031184.
21. Rietschel L, Zhu G, Kirschbaum C, Strohmaier J, Wust S, Rietschel M, et al. Perceived stress has genetic influences distinct from neuroticism and depression. Behav Genet. 2014; 44(6):639–45. https://doi.org/10.1007/s10519-013-9636-4 PMID: 24366676.

22. Rietschel L, Streit F, Zhu G, McAloney K, Frank J, Couvy-Duchesne B, et al. Hair Cortisol in Twins: Heritability and Genetic Overlap with Psychological Variables and Stress-System Genes. Sci Rep. 2017; 7(1):15351. https://doi.org/10.1038/s41598-017-11852-3 PMID: 29127340; PubMed Central PMCID: PMCPMC5703444.

23. Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr., Charlesworth JC, et al. High dimensional endophenotype ranking in the search for major depression risk genes. Biol Psychiatry. 2012; 71(1):6–14. https://doi.org/10.1016/j.biopsych.2011.08.022 PMID: 21982424; PubMed Central PMCID: PMCPMC3230692.

24. van den Berg SM, de Moor MH, McGue M, Pettersson E, Terracciano A, Verweij KJ, et al. Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. Behav Genet. 2014; 44(4):295–313. https://doi.org/10.1007/s10519-014-9654-x PMID: 24828478; PubMed Central PMCID: PMCPMC4057636.

25. Wray NR, Birley AJ, Sullivan PF, Visscher PM, Martin NG. Genetic and phenotypic stability of measures of neuroticism over 22 years. Twin Res Hum Genet. 2007; 10(5):695–702. https://doi.org/10.1375/twin.10.5.695 PMID: 17903109.

26. Vukasovic T, Bratko D. Heritability of personality: A meta-analysis of behavior genetic studies. Psychol Bull. 2015; 141(4):769–85. https://doi.org/10.1037/bul0000017 PMID: 25961374.

27. Hirschfeld RM, Klerman GL. Personality attributes and affective disorders. Am J Psychiatry. 1979; 136 (1):67–70. https://doi.org/10.1176/ajp.136.1.67 PMID: 758831.

28. Hirschfeld RM, Klerman GL, Clayton PJ, Keller MB, McDonald-Scott P, Larkin BH. Assessing personality: effects of the depressive state on trait measurement. Am J Psychiatry. 1983; 140(6):695–9. https://doi.org/10.1176/ajp.140.6.695 PMID: 6846626.

29. Levinson DF. The genetics of depression: a review. Biol Psychiatry. 2006; 60(2):84–92. https://doi.org/10.1016/j.biopsych.2005.08.024 PMID: 16300747.

30. Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. Psychol Med. 2005; 35 (5):611–24. PMID: 15918338.

31. Smith DJ, Escott-Price V, Davies G, Bailey ME, Colodro-Conde L, Ward J, et al. Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. Mol Psychiatry. 2016; 21 (6):749–57. https://doi.org/10.1038/mp.2016.49 PMID: 27067015; PubMed Central PMCID: PMCPMC4879189.

32. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018; 50 (1):6–11. https://doi.org/10.1038/s41588-017-0013-8 PMID: 29255261.

33. Smits DJM, Boeck PD. From BIS/BAS to the big five. European Journal of Personality. 2006; 20 (4):255–70. https://doi.org/10.1002/per.583

34. Schneider TR, Rench TA, Lyons JB, Riffle RR. The influence of neuroticism, extraversion and openness on stress responses. Stress Health. 2012; 28(2):102–10. https://doi.org/10.1002/smi.1409 PMID: 22281953.

35. Kim SE, Kim HN, Cho J, Kwon MJ, Chang Y, Ryu S, et al. Direct and Indirect Effects of Five Factor Personality and Gender on Depressive Symptoms Mediated by Perceived Stress. PLoS One. 2016; 11(4):e0154140. https://doi.org/10.1371/journal.pone.0154140 PMID: 27120051; PubMed Central PMCID: PMCPMC4847785.

36. Tak LM, Kingma EM, van Ockenburg SL, Ormel J, Rosmalen JG. Age- and sex-specific associations between adverse life events and functional bodily symptoms in the general population. J Psychosom Res. 2015; 79(2):112–6. https://doi.org/10.1016/j.jpsychores.2015.05.013 PMID: 26052060.

37. Hovens JG, Giltay EJ, van Hemert AM, Penninx BW. Childhood Maltreatment and the Course of Depressive and Anxiety Disorders: The Contribution of Personality Characteristics. Depress Anxiety. 2016; 33(1):27–34. https://doi.org/10.1002/da.22429 PMID: 26418232.

38. Bazana PG, Stelmack RM. Chapter 8—Stability of Personality Across the Life Span: A Meta-Analysis. On the Psychobiology of Personality. Oxford: Elsevier; 2004. p. 113–44.

39. McGue M, Bacon S, Lykken DT. Personality Stability and Change in Early Adulthood. Developmental Psychology. 1993; 29(1):96–109.

40. Nivard MG, Middeldorp CM, Dolan CV, Boomsma DI. Genetic and Environmental Stability of Neuroticism From Adolescence to Adulthood. Twin Res Hum Genet. 2015; 18(6):746–54. https://doi.org/10.1017/thg.2015.80 PMID: 26678053.

41. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000; 157(10):1552–62. https://doi.org/10.1176/appi.ajp.157.10.1552 PMID: 11007705.

42. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018; 50(5):668–81. https://doi.org/10.1038/s41588-018-0090-3 PMID: 29700475; PubMed Central PMCID: PMCPMC5934326.

43. Loret de Mola C, de Franca GV, Quevedo Lde A, Horta BL. Low birth weight, preterm birth and small for gestational age association with adult depression: systematic review and meta-analysis. Br J Psychiatry. 2014; 205(5):340–7. https://doi.org/10.1192/bjp.bp.113.139014 PMID: 25368358.

44. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. Br J Psychiatry. 2014; 205(2):113–9. https://doi.org/10.1192/bjp.bp.113.143081 PMID: 24925986; PubMed Central PMCID: PMCPMC4118052.

45. Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychol Med. 2016; 46 (4):759–70. https://doi.org/10.1017/S0033291715002172 PMID: 26526099; PubMed Central PMCID: PMCPMC4754832.

46. Colodro-Conde L, Couvy-Duchesne B, Zhu G, Coventry WL, Byrne EM, Gordon S, et al. A direct test of the diathesis-stress model for depression. Mol Psychiatry. 2017. https://doi.org/10.1038/mp.2017.130 PMID: 28696435.

47. Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. Psychol Med. 2007; 37(5):615–26. https://doi.org/10.1017/S0033291706009524 PMID: 17176502.

48. Luciano M, Batty GD, McGilchrist M, Linksted P, Fitzpatrick B, Jackson C, et al. Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. Intelligence. 2010; 38(3):304–13. http://dx.doi.org/10.1016/j.intell.2010.03.002.

49. Silberg J, Rutter M, Neale M, Eaves L. Genetic moderation of environmental risk for depression and anxiety in adolescent girls. Br J Psychiatry. 2001; 179:116–21. PMID: 11483472.

50. Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. Am J Psychiatry. 1995; 152 (6):833–42. https://doi.org/10.1176/ajp.152.6.833 PMID: 7755111.

51. Biobank U. UK Biobank: protocol for a large-scale prospective epidemiological resource 2010. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf.

52. Biobank U. UK Biobank ethics and governance framework, version 3.0 2007. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6.

53. Biobank U. Genotype imputation and genetic association studies of UK Biobank. 2015.

54. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81 (3):559–75. https://doi.org/10.1086/519795 PMID: 17701901; PubMed Central PMCID: PMCPMC1950838.

55. Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Generation S, et al. Recent genomic heritage in Scotland. BMC Genomics. 2015; 16:437. https://doi.org/10.1186/s12864-015-1605-2 PMID: 26048416; PubMed Central PMCID: PMCPMC4458001.

56. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013; 42(3):689–700. https://doi.org/10.1093/ije/dys084 PMID: 22786799.

57. Fernandez-Pujals AM, Adams MJ, Thomson P, McKechanie AG, Blackwood DH, Smith BH, et al. Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). PLoS One. 2015; 10(11): e0142197. https://doi.org/10.1371/journal.pone.0142197 PMID: 26571028; PubMed Central PMCID: PMCPMC4646689.

58. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC medical genetics. 2006; 7:74. https://doi.org/10.1186/1471-2350-7-74 PMID: 17014726; PubMed Central PMCID: PMC1592477.

59. Kerr SM, Campbell A, Murphy L, Hayward C, Jackson C, Wain LV, et al. Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. BMC Med Genet. 2013; 14:38. https://doi.org/10.1186/1471-2350-14-38 PMID: 23521772; PubMed Central PMCID: PMCPMC3614907.

60. Biobank U. Touchscreen questionnaire 2012. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Touch_screen_questionnaire.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6.

61. Eysenck HJ ES. Manual of the Eysenck Personality Questionnaire. London: Hodder and Stoughton; 1975.

62. Kendler KS, Myers J. The genetic and environmental relationship between major depression and the five-factor model of personality. Psychol Med. 2010; 40(5):801–6. https://doi.org/10.1017/S0033291709991140 PMID: 19732485.

63. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. PLoS One. 2013; 8(11):e75362. https://doi.org/10.1371/journal.pone.0075362 PMID: 24282498; PubMed Central PMCID: PMCPMC3839907.

64. Eysenck SBG, Eysenck H.J. & Barrett P. A revised version of the psychoticism scale. Personality and Individual Differences. 1985; 6:21–9.

65. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA. 1999; 282(18):1737–44. PMID: 10568646.

66. First MB SR, Gibbon M, Williams JB. Structured Clinical Interview for DSM-IV-TR Axis I Disorders. New York: New York State Psychiatric Institute; 2002.

67. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015; 31 (9):1466–8. https://doi.org/10.1093/bioinformatics/btu848 PMID: 25550326; PubMed Central PMCID: PMCPMC4410663.

68. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26(17):2190–1. https://doi.org/10.1093/bioinformatics/btq340 PMID: 20616382; PubMed Central PMCID: PMCPMC2922887.

69. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015; 11(4):e1004219. https://doi.org/10.1371/journal.pcbi.1004219 PMID: 25885710; PubMed Central PMCID: PMCPMC4401657.

70. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. FUMA: Functional mapping and annotation of genetic associations. bioRxiv. 2017. https://doi.org/10.1101/110023

71. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47(3):291–5. https://doi.org/10.1038/ng.3211 PMID: 25642630; PubMed Central PMCID: PMCPMC4495769.

72. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015; 47(11):1236–41. https://doi.org/10.1038/ng.3406 PMID: 26414676; PubMed Central PMCID: PMCPMC4797329.

73. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. 2015; 6:5890. https://doi.org/10.1038/ncomms6890 PMID: 25597830; PubMed Central PMCID: PMCPMC4420238.

74. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39(10):1181–6. https://doi.org/10.1038/ng1007-1181 PMID: 17898773; PubMed Central PMCID: PMCPMC2031016.

75. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012; 22(9):1790–7. https://doi.org/10.1101/gr.137323.112 PMID: 22955989; PubMed Central PMCID: PMC3431494.

76. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348(6235):648–60. https://doi.org/10.1126/science.1262110 PMID: 25954001; PubMed Central PMCID: PMCPMC4547484.

77. Genetics of Personality C, de Moor MH, van den Berg SM, Verweij KJ, Krueger RF, Luciano M, et al. Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. JAMA Psychiatry. 2015; 72(7):642–50. https://doi.org/10.1001/jamapsychiatry.2015.0554 PMID: 25993607; PubMed Central PMCID: PMCPMC4667957.

78. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

79. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011; 88(3):294–305. https://doi.org/10.1016/j.ajhg.2011.02.002 PMID: 21376301; PubMed Central PMCID: PMCPMC3059431.

80. Chen GB. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. Front Genet. 2014; 5:107. https://doi.org/10.3389/fgene.2014.00107 PMID: 24817879; PubMed Central PMCID: PMCPMC4012219.

81. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE Jr., et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am J Psychiatry. 2000; 157(11):1873–5. https://doi.org/10.1176/appi.ajp.157.11.1873 PMID: 11058490.

82. Raine A. The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. Schizophr Bull. 1991; 17(4):555–64. PMID: 1805349.

83. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015; 47(7):702–9. https://doi.org/10.1038/ng.3285 PMID: 25985137.

84. consortium C. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015; 523(7562):588–91. https://doi.org/10.1038/nature14659 PMID: 26176920; PubMed Central PMCID: PMCPMC4522619.

85. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016; 48(9):1031–6. https://doi.org/10.1038/ng.3623 PMID: 27479909.

86. Zuo L, Tan Y, Zhang X, Wang X, Krystal J, Tabakoff B, et al. A New Genomewide Association Meta-Analysis of Alcohol Dependence. Alcohol Clin Exp Res. 2015; 39(8):1388–95. https://doi.org/10.1111/acer.12786 PMID: 26173551.

87. Cook MN, Baker JA, Heldt SA, Williams RW, Hamre KM, Lu L. Identification of candidate genes that underlie the QTL on chromosome 1 that mediates genetic differences in stress-ethanol interactions. Physiol Genomics. 2015; 47(8):308–17. https://doi.org/10.1152/physiolgenomics.00114.2014 PMID: 25991709; PubMed Central PMCID: PMCPMC4525077.

88. Zuo L, Wang K, Wang G, Pan X, Zhang X, Zhang H, et al. Common PTP4A1-PHF3-EYS variants are specific for alcohol dependence. Am J Addict. 2014; 23(4):411–4. https://doi.org/10.1111/j.1521-0391.2013.12115.x PMID: 24961364; PubMed Central PMCID: PMCPMC4111256.

89. Zuo L, Zhang X, Deng HW, Luo X. Association of rare PTP4A1-PHF3-EYS variants with alcohol dependence. J Hum Genet. 2013; 58(3):178–9. https://doi.org/10.1038/jhg.2012.153 PMID: 23324950.

90. Zuo L, Zhang CK, Wang F, Li CS, Zhao H, Lu L, et al. A novel, functional and replicable risk gene region for alcohol dependence identified by genome-wide association study. PLoS One. 2011; 6(11):e26726. https://doi.org/10.1371/journal.pone.0026726 PMID: 22096494; PubMed Central PMCID: PMCPMC3210123.

91. Hontelez S, Karthaus N, Looman MW, Ansems M, Adema GJ. DC-SCRIPT regulates glucocorticoid receptor function and expression of its target GILZ in dendritic cells. J Immunol. 2013; 190(7):3172–9. https://doi.org/10.4049/jimmunol.1201776 PMID: 23440419.

92. Lopez-Garcia J, Periyasamy M, Thomas RS, Christian M, Leao M, Jat P, et al. ZNF366 is an estrogen receptor corepressor that acts through CtBP and histone deacetylases. Nucleic Acids Res. 2006; 34(21):6126–36. https://doi.org/10.1093/nar/gkl875 PMID: 17085477; PubMed Central PMCID: PMCPMC1693901.

93. Miyamoto-Sato E, Fujimori S, Ishizaka M, Hirai N, Masuoka K, Saito R, et al. A comprehensive resource of interacting protein regions for refining human transcription factor networks. PLoS One. 2010; 5(2):e9289. https://doi.org/10.1371/journal.pone.0009289 PMID: 20195357; PubMed Central PMCID: PMCPMC2827538.

94. Aragam N, Wang KS, Pan Y. Genome-wide association analysis of gender differences in major depressive disorder in the Netherlands NESDA and NTR population-based samples. J Affect Disord. 2011; 133(3):516–21. https://doi.org/10.1016/j.jad.2011.04.054 PMID: 21621269.

95. Fidalgo TM, da Silveira ED, da Silveira DX. Psychiatric comorbidity related to alcohol use among adolescents. Am J Drug Alcohol Abuse. 2008; 34(1):83–9. https://doi.org/10.1080/00952990701764664 PMID: 18161646.

96. Lopez B, Turner RJ, Saavedra LM. Anxiety and risk for substance dependence among late adolescents/young adults. J Anxiety Disord. 2005; 19(3):275–94. https://doi.org/10.1016/j.janxdis.2004.03.001 PMID: 15686857.

97. Low NC, Lee SS, Johnson JG, Williams JB, Harris ES. The association between anxiety and alcohol versus cannabis abuse disorders among adolescents in primary care settings. Fam Pract. 2008; 25(5):321–7. https://doi.org/10.1093/fampra/cmn049 PMID: 18753288.

98. Rutledge PC, Sher KJ. Heavy drinking from the freshman year into early young adulthood: the roles of stress, tension-reduction drinking motives, gender and personality. J Stud Alcohol. 2001; 62(4):457–66. PMID: 11523533.

99. Schmidt NB, Buckner JD, Keough ME. Anxiety sensitivity as a prospective predictor of alcohol use disorders. Behav Modif. 2007; 31(2):202–19. https://doi.org/10.1177/0145445506297019 PMID: 17307935.

100. Varlinskaya EI, Kim EU, Spear LP. Chronic intermittent ethanol exposure during adolescence: Effects on stress-induced social alterations and social drinking in adulthood. Brain Res. 2017; 1654(Pt B):145–56. https://doi.org/10.1016/j.brainres.2016.03.050 PMID: 27048754; PubMed Central PMCID: PMCPMC5047849.

101. Torigata K, Daisuke O, Mukai S, Hatanaka A, Ohka F, Motooka D, et al. LATS2 Positively Regulates Polycomb Repressive Complex 2. PLoS One. 2016; 11(7):e0158562. https://doi.org/10.1371/journal.pone.0158562 PMID: 27434182; PubMed Central PMCID: PMCPMC4951031.

102. Shen EY, Jiang Y, Javidfar B, Kassim B, Loh YE, Ma Q, et al. Neuronal Deletion of Kmt2a/Mll1 Histone Methyltransferase in Ventral Striatum is Associated with Defective Spike-Timing-Dependent Striatal Synaptic Plasticity, Altered Response to Dopaminergic Drugs, and Increased Anxiety. Neuropsychopharmacology. 2016; 41(13):3103–13. https://doi.org/10.1038/npp.2016.144 PMID: 27485686; PubMed Central PMCID: PMCPMC5101561.

103. Wong ML, Arcos-Burgos M, Liu S, Velez JI, Yu C, Baune BT, et al. The PHF21B gene is associated with major depression and modulates the stress response. Mol Psychiatry. 2017; 22(7):1015–25. https://doi.org/10.1038/mp.2016.174 PMID: 27777418; PubMed Central PMCID: PMCPMC5461220.

104. Walsh RM, Shen EY, Bagot RC, Anselmo A, Jiang Y, Javidfar B, et al. Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. Nat Commun. 2017; 8:15142. https://doi.org/10.1038/ncomms15142 PMID: 28485378; PubMed Central PMCID: PMCPMC5436068.

105. Gage SH, Jones HJ, Taylor AE, Burgess S, Zammit S, Munafo MR. Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomization. Sci Rep. 2017; 7:40653. https://doi.org/10.1038/srep40653 PMID: 28102331; PubMed Central PMCID: PMCPMC5244403.

106. Hartz SM, Horton AC, Hancock DB, Baker TB, Caporaso NE, Chen LS, et al. Genetic correlation between smoking behaviors and schizophrenia. Schizophr Res. 2017. https://doi.org/10.1016/j.schres.2017.02.022 PMID: 28285025.

107. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9(3): e1003348. https://doi.org/10.1371/journal.pgen.1003348 PMID: 23555274; PubMed Central PMCID: PMCPMC3605113.

Appendix B

Appendix B contains supplementary material for **chapter 3**: Enrichment of genetic variation conferring MDD risk in glucocorticoid-related genesets: partitioning risk based on main additive contributions to MDD, neuroticism and the proxy for stress-sensitivity.

B.1 Supplementary Tables

Supplementary Table 1 List of genes from each geneset.

List of genes geneset 1. In bold overlapping genes with geneset 2 (**blue**), with geneset 3 (**green**), and with both geneset 2 and geneset 3 (**red**).

A2M	**ATP1A1**	**CCNE1**	**EGFR**	HDAC3	**KCNMA1**
AANAT	ATP2B1	**CDKN1A**	**ELN**	HDAC6	**KLF9**
ABCA3	AVPR1A	CDO1	EP300	HEY1	KRAS
ABCB1	BAD	CEBPA	EPHX1	HIF1A	LCAT
ABCB4	BAX	**CEBPB**	**ERRFI1**	HLA-DQA1	**LIPC**
ABCG2	BCHE	CHRFAM7A	ESR2	HLA-DQB1	**LMO3**
ACADM	**BCKDHA**	CLCF1	**ETS2**	HLA-DQB2	MAOB
ACADS	BCKDHB	CLOCK	FABP4	HMBS	**MAP2K1**
ACE	BCL2	CPN1	FAS	HMGCS2	MC2R
ACSBG1	**BDH1**	CPNE1	**FBXO32**	**HNMT**	MDK
ACTG2	**BDNF**	**CPS1**	FECH	HNRNPA2B1	**MECP2**
ACVR1	**BGLAP**	CREBRF	FKBP4	HP	MGP
ADAM9	BIRC5	CRH	**FLT3**	HPR	MIF
ADCYAP1	BMI1	CRHBP	**FN1**	**HSD11B1**	MSTN
ADIPOQ	BMP2	CRHR1	FOS	**HSD11B2**	**MYOC**
ADM	BMP4	CRHR2	FOSB	**HSD17B3**	MYOD1
ADRB2	**BMP5**	**CRY1**	FOSL1	HSD3B1	**NCOA2**
AGL	**BMP6**	**CRY2**	**FOSL2**	HSD3B2	NCOA6
AGTR1	C1QB	**CRYAA**	**GAL**	**HTR1B**	**NCOR2**
AGTR2	**C1QL1**	CSDC2	**GALR1**	HTR2C	**NEDD4**
AGXT	**C1QTNF1**	**CTGF**	GALT	HTR7	**NEFL**
AHR	**C3**	CTSV	GBA	IFNB1	NGF
AIF1	C8orf44-SGK3	CXCL1	GDNF	IFNLR1	**NOS2**
AKAP13	**CACNA1G**	**CXCL2**	GHR	**IGF1**	NOS3
ALAD	**CACNA1H**	CXCL3	**GHRHR**	**IGFBP2**	NPAS4
ALDH3A1	CAD	CYP11A1	GHRL	**IGFBP7**	**NR3C1**
ALDOB	CALCR	CYP11B1	GLCCI1	IL10	NR3C2
ALPL	**CALM1**	CYP11B2	GMEB1	IL10RB	**NRIP1**
ANXA1	**CALM2**	CYP17A1	GMEB2	IL17A	**NTRK2**
ANXA3	**CALM3**	**CYP1B1**	**GNAL**	IL1B	**NTRK3**
APOA1	CARHSP1	**CYP21A2**	**GNAO1**	IL1RN	ORM1
APOA2	CASP3	**CYP3A4**	GNAS	IL22	ORM2
AQP1	CASP9	**CYP3A5**	**GOT1**	IL6	OXT
AQP4	**CAV1**	**CYP3A7**	GPER1	**IL6R**	PAM
AREG	**CBX3**	**DAB2**	GPR83	**INSR**	**PAPPA**
ARG1	CCKAR	**DKK3**	GPX3	IPO13	**PARK2**
ARG2	CCL1	DNMT3B	**GRIP1**	**ISL1**	PCK2
ARID1A	**CCL2**	**DPYD**	GSTP1	JAK2	**PCSK1**
ARNTL	CCL5	**DUSP1**	GTPBP4	**JUNB**	PDCD7
ASS1	**CCND1**	**EDN1**	H19	KCNJ10	PDX1

PEBP1	PTGES3	SERPINF1	SLIT3	SULT1A4	UCN2
PER1	PTGS1	SFRP4	SMYD3	TAC1	UCN3
PFKFB1	PTGS2	SFTPA1	SOCS3	TAT	UCP3
PHB	PTPN11	SFTPA2	SPARC	TFAP4	UGT1A1
PIAS2	PTPRU	SFTPB	SRC	TGFB1	UGT1A5
PIK3R1	RBM14	SFTPC	SRD5A1	TH	UGT1A6
PLA2G4A	RELN	SFTPD	SSTR2	TK1	UGT2B10
PLAT	REST	SGK1	SSTR3	TLR4	UGT2B11
PNLIPRP1	RHOA	SGK2	SSTR4	TMPRSS4-AS1	UGT2B15
PNLIPRP2	RPS6KB1	SGK3	SSTR5	TNF	UGT2B17
PNMT	RXRA	SHC1	STAR	TNFSF11	UGT2B28
POMC	S100B	SI	STAT3	TPH2	UGT2B4
PPARGC1B	SCGB1A1	SLC18A2	STAT5B	TRH	UGT2B7
PPP5C	SDC1	SLC9A3	STC1	TRIM63	VEGFA
PRKCA	SERPINA3	SLCO1B1	STUB1	TXN	WNT4
PRSS8	SERPINA6	SLCO1B3	SULT1A1	TYMS	WNT5A
PSMC3IP	SERPINA7	SLCO1B7	SULT1A2	UBE2L3	WNT7B
PTGDS	SERPINE1	SLIT2	SULT1A3	UCN	YWHAH

List of genes geneset 2

A4GALT	AC005062.2	ACMSD	ADCK4	AGR3	ALOX15B	ANKRD42
AACS	AC006196.1	ACOT1	ADCY1	AGRP	ALOX15P1	ANKRD44
AACSP1	AC006372.4	ACOT11	ADCY10	AGTPBP1	ALOX5AP	ANKRD44-IT1
AAMDC	AC007163.6	ACOT2	ADCY3	AGTR2	ALPK2	ANKRD46
AATK	AC007246.3	ACOT7	ADCY8	AHCYL1	ALPL	ANKRD55
ABCA1	AC007879.2	ACOX2	ADCY9	AHDC1	ALPP	ANKS1A
ABCA13	AC007880.1	ACOXL	ADD1	AHNAK	AMBP	ANKS1B
ABCA3	AC008697.1	ACPL2	ADD2	AHNAK2	AMFR	ANKS6
ABCA4	AC008703.1	ACRC	ADD3	AHRR	AMMECR1	ANO1
ABCA7	AC010987.6	ACSBG1	ADI1	AHSA2	AMN1	ANO1-AS2
ABCB11	AC011343.1	ACSF2	ADK	AICDA	AMOTL1	ANO2
ABCB5	AC011747.4	ACSF3	ADM	AIG1	AMOTL2	ANO6
ABCB8	AC011747.7	ACSL1	ADNP	AIM1L	AMPD3	ANPEP
ABCB9	AC011899.9	ACSL3	ADNP2	AK4	AMZ1	ANTXR1
ABCC1	AC016995.3	ACSL4	ADORA2A	AK5	AMZ2	ANXA1
ABCC11	AC017060.1	ACSM5	ADORA2B	AK8	AMZ2P1	ANXA13
ABCC12	AC018470.4	ACSS1	ADORA3	AKAP10	ANAPC16	ANXA2
ABCC2	AC018705.5	ACTB	ADPRHL1	AKAP13	ANG	ANXA2R
ABCC3	AC019048.1	ACTG1	ADRA1B	AKAP2	ANGPT1	ANXA4
ABCC4	AC019117.1	ACTL8	ADRA2A	AKAP6	ANGPT4	ANXA5
ABCC5	AC020571.3	ACTN1	ADRA2B	AKIRIN1	ANGPTL4	ANXA8L1
ABCC6	AC022173.2	ACTN4	ADRA2C	AKNA	ANK3	AOC2
ABCC8	AC073283.4	ACTR10	ADSSL1	AKNAD1	ANKDD1A	AOX1
ABCC9	AC073321.3	ACTR3BP2	ADTRP	AKR1B10	ANKDD1B	AOX2P
ABCD2	AC074286.1	ACTR3BP5	AEBP2	AKR1B15	ANKFN1	AP1M1
ABCD3	AC079610.2	ACTRT3	AEN	AKR1C1	ANKFY1	AP1S2
ABCF1	AC090044.1	ACVR1	AES	AKR1C2	ANKH	AP1S3
ABCG1	AC090505.6	ACVR1B	AFAP1	AKR1C3	ANKLE2	AP2A1
ABCG2	AC091878.1	ACYP2	AFAP1L1	AKR1CL1	ANKRA2	AP2A2
ABCG5	AC093590.1	ADAD2	AFAP1L2	AKR1D1	ANKRD1	AP2B1
ABHD12	AC093627.10	ADAM17	AFF1	AKR1E2	ANKRD11	AP3B1
ABHD16B	AC095067.1	ADAM28	AFF4	AKT3	ANKRD13A	AP3S2
ABHD17C	AC096559.1	ADAM8	AGAP1	ALCAM	ANKRD2	AP4M1
ABHD2	AC097662.2	ADAM9	AGAP11	ALDH1A1	ANKRD20A11P	AP5B1
ABHD3	AC098823.3	ADAMTS10	AGAP5	ALDH1A2	ANKRD20A2	AP5Z1
ABHD4	AC104654.2	ADAMTS12	AGAP8	ALDH1L1-AS2	ANKRD20A3	APBA1
ABHD5	AC104777.2	ADAMTS15	AGBL1	ALDH2	ANKRD20A5P	APBB1IP
ABHD6	AC104777.4	ADAMTS16	AGBL4	ALDH3A1	ANKRD20A8P	APBB2
ABI3BP	AC105053.3	ADAMTS18	AGFG1	ALDH3A2	ANKRD20A9P	APCS
ABL1	AC107070.1	ADAMTS7	AGFG2	ALDH3B1	ANKRD22	APIP
ABL2	AC125421.1	ADAMTS8	AGGF1	ALDH3B2	ANKRD26P1	APITD1
ABLIM1	AC144449.1	ADAMTS9	AGMAT	ALDH7A1	ANKRD26P3	APITD1-CORT
ABLIM2	ACACA	ADAMTS9-AS2	AGMO	ALDH8A1	ANKRD27	APOA4
ABLIM3	ACACB	ADAMTSL1	AGPAT1	ALDOB	ANKRD28	APOD
ABR	ACAN	ADAMTSL4	AGPAT3	ALG14	ANKRD29	APOH
ABTB2	ACBD7	ADAMTSL4-AS1	AGPAT4	ALG1L9P	ANKRD30A	APOL1
AC003051.1	ACE	ADARB1	AGPAT6	ALG8	ANKRD30BP2	APOL3
AC003984.1	ACER3	ADCK1	AGPAT9	ALG9	ANKRD33B	APOO
AC004603.4	ACLY	ADCK3	AGR2	ALKBH3-AS1	ANKRD40	APP

APPL1
APPL2
APRT
AQP3
AQP7
AQP7P1
AQP7P3
AR
ARAF
ARAP2
AREG
ARFGAP3
ARFGEF2
ARFRP1
ARHGAP10
ARHGAP11B
ARHGAP12
ARHGAP17
ARHGAP18
ARHGAP19
ARHGAP23
ARHGAP24
ARHGAP25
ARHGAP26
ARHGAP26-AS1
ARHGAP27
ARHGAP29
ARHGAP31
ARHGAP32
ARHGAP35
ARHGAP40
ARHGAP42
ARHGAP5-AS1
ARHGAP6
ARHGAP8
ARHGEF10L
ARHGEF12
ARHGEF18
ARHGEF19
ARHGEF26
ARHGEF26-AS1
ARHGEF28
ARHGEF3
ARHGEF3-AS1
ARHGEF37
ARHGEF38
ARHGEF39
ARHGEF7
ARID1B
ARID5B
ARL13A
ARL14
ARL15
ARL17B
ARL3
ARL4A
ARL4D
ARL5A
ARL8B
ARMC5
ARMCX3
ARNT2
ARNTL
ARPP21
ARRB1
ARRB2
ARRDC1
ARRDC2
ARRDC3
ARRDC3-AS1
ARSB
ARSD
ARSG
ARSI
ART1
ART4
ARX
ASAH2B
ASAP1
ASAP2
ASAP3
ASB1
ASB13
ASB4
ASB7
ASB9
ASB9P1
ASCC1
ASCC2
ASCC3
ASIC2
ASIP
ASL
ASMT
ASNA1
ASPH
ASRGL1
ASS1
ASTL
ASTN2
ATAD2
ATAD2B
ATAT1
ATE1
ATF1
ATF3
ATF6
ATF7
ATG10
ATG13
ATG16L1
ATG5
ATG7
ATOH8
ATP10A
ATP11A
ATP11AUN
ATP11B
ATP13A4
ATP1A1
ATP1A2
ATP1A4
ATP1B1
ATP1B3
ATP2A2
ATP2B2
ATP2B4
ATP2C2
ATP4A
ATP5D
ATP5G1
ATP5G2
ATP5S
ATP6V0D1
ATP6V0E1
ATP6V1G3
ATP7A
ATP7B
ATP8B1
ATP8B4
ATP8B5P
ATP9A
ATPAF1
ATPAF2
ATRN
ATRX
ATXN1
ATXN10
ATXN1L
ATXN2
ATXN7
ATXN7L1
AUH
AUTS2
AVEN
AVL9
AVPI1
AWAT1
AXIN2
AZGP1P1
AZU1
B3GALNT2
B3GALT1
B3GALT5
B3GNT2
B3GNT6
B4GALNT2
B4GALT1
B4GALT4
B4GALT5
B9D1
BAAT
BACH2
BAG3
BAGE
BAGE2
BAGE5
BAHCC1
BAIAP2
BAIAP2L1
BAMBI
BANK1
BANP
BAP1
BATF
BATF3
BBOX1
BBS10
BBS12
BBS2
BBX
BCAN
BCAR1
BCAR3
BCAR4
BCAS1
BCAS3
BCAS4
BCAT1
BCKDHA
BCL2A1
BCL2L1
BCL2L11
BCL2L14
BCL3
BCL6
BCL7A
BCL7C
BCL9L
BCMS
BCOR
BCR
BCRP3
BDH1
BDKRB2
BDNF
BDP1
BEGAIN
BEND2
BEST1
BEST3
BEST4
BFAR
BFSP1
BFSP2
BGLAP
BHLHE40
BHLHE40-AS1
BICC1
BICD1
BIN1
BIN2
BIRC2
BIRC3
BIRC6
BLK
BLM
BLOC1S1
BLOC1S4
BLOC1S5
BLVRA
BMF
BMP1
BMP5
BMP6
BMP8A
BMPER
BMPR1B
BMPR2
BMS1P4
BNC1
BNC2
BNIP2
BOK-AS1
BOLA2
BOLA2B
BPESC1
BPIFA4P
BPTF
BRCA1
BRD3
BRD4
BRE
BRF1
BRINP2
BRINP3
BRSK2
BSN
BTAF1
BTBD1
BTBD11
BTBD16
BTBD9
BTC
BTD
BTNL9
BTRC
BUB1B
BZW2
C10orf107
C10orf11
C10orf113
C10orf12
C10orf40
C10orf54
C10orf76
C10orf90
C11orf35
C11orf49
C11orf54
C11orf63
C11orf70
C11orf71
C11orf80
C11orf86
C11orf91
C12orf36
C12orf61
C12orf75
C12orf76
C12orf79
C14orf132
C14orf159
C14orf164
C14orf182
C14orf37
C15orf27
C15orf32
C15orf38
C15orf38-AP3S2
C15orf39
C15orf41
C15orf54
C15orf59
C16orf11
C16orf45
C16orf52
C16orf72
C16orf74
C16orf96
C17orf103
C17orf105
C17orf51
C17orf53
C17orf58
C17orf67
C17orf82
C18orf25
C18orf54
C18orf61
C19orf35
C19orf40
C1GALT1C1
C1QBP
C1QL1
C1QL3
C1QTNF1
C1QTNF1-AS1
C1QTNF6
C1QTNF8
C1QTNF9
C1R
C1RL
C1RL-AS1
C1S
C1orf100
C1orf105
C1orf106
C1orf116
C1orf168
C1orf172
C1orf180
C1orf192
C1orf195
C1orf204
C1orf21
C1orf213
C1orf226
C1orf228
C1orf234
C1orf27
C1orf86
C2
C20orf112
C20orf194
C20orf196
C20orf203
C20orf26
C20orf62
C20orf85
C21orf33
C22orf26
C2CD2
C2CD3
C2CD4A
C2CD4B
C2orf47
C2orf57
C2orf61
C2orf72
C2orf73
C2orf76
C2orf80
C2orf83
C2orf91
C3
C3orf20
C3orf43
C4orf45
C4orf46
C4orf51
C5
C5AR1
C5AR2
C5orf27
C5orf28
C5orf30
C5orf34
C5orf47
C5orf63
C5orf64
C5orf66
C5orf66-AS2
C6
C6orf106
C6orf132
C6orf195
C6orf201
C6orf211
C6orf223
C6orf57
C6orf89
C6orf99
C7orf43
C7orf49
C7orf50
C7orf65
C7orf69
C7orf71
C8orf31
C8orf34
C8orf4
C8orf46
C8orf74
C8orf76
C8orf86
C9orf106
C9orf116
C9orf117
C9orf129
C9orf135
C9orf3
C9orf62
C9orf69
C9orf72
C9orf89
C9orf91
CA12
CA4
CA5B
CA5BP1
CA8
CA9
CAB39L
CABIN1
CABLES1
CABLES2
CACHD1
CACNA1C
CACNA1C-AS1
CACNA1D
CACNA1G
CACNA1H
CACNA1I
CACNA2D1
CACNA2D3
CACNA2D3-AS1
CACNB4
CACNG1
CACNG3
CACNG6
CADM1
CADM2
CAGE1
CALCOCO2
CALD1
CALM1
CALM2
CALM3
CALML3
CALML5
CALN1
CAMK1D
CAMK1G
CAMK2A
CAMK2D
CAMK2G
CAMK4
CAMKK1
CAMKMT
CAMTA1
CAND1.11
CAP2
CAPN12
CAPN2
CAPN7
CAPN8
CAPZA1
CAPZA2
CARD10
CARD11
CARD14
CARTPT
CASC15
CASC16
CASC17
CASC18
CASC2
CASC8
CASK
CASQ2
CASR
CAST
CASZ1

CATIP
CATSPERB
CATSPERD
CAV1
CAV2
CBFA2T3
CBLB
CBX3
CBX4
CBX7
CCAT1
CCBE1
CCBL1
CCDC101
CCDC102A
CCDC109B
CCDC112
CCDC12
CCDC137
CCDC141
CCDC144A
CCDC144B
CCDC144CP
CCDC144NL
CCDC158
CCDC26
CCDC33
CCDC37-AS1
CCDC40
CCDC57
CCDC62
CCDC63
CCDC64
CCDC68
CCDC77
CCDC80
CCDC85A
CCDC85C
CCDC88C
CCDC91
CCDC97
CCL2
CCL20
CCND1
CCND3
CCNE1
CCNG1
CCNH
CCNL1
CCNY
CCNYL1
CCPG1
CCR7
CCR9
CCRN4L
CCSER2
CCT4
CCZ1B
CD109
CD151
CD180
CD200R1L
CD247
CD27-AS1
CD28
CD2AP
CD300C
CD300E
CD300LG
CD36
CD38
CD44
CD53
CD55
CD58
CD59
CD70
CD82
CD8B
CD96
CD99
CD99L2
CD99P1
CDC123
CDC14B
CDC16
CDC27
CDC42BPA
CDC42BPB
CDC42EP2
CDC42EP3
CDC42EP4
CDC42SE2
CDC5L
CDC7
CDCA2
CDCA7
CDCA7L
CDCP2
CDH16
CDH17
CDH18
CDH2
CDH22
CDH23
CDH3
CDH5
CDHR2
CDK1
CDK19
CDK3
CDK5R2
CDK5RAP1
CDK5RAP2
CDK5RAP3
CDK6
CDKAL1
CDKL1
CDKL3
CDKL5
CDKN1A
CDKN2B-AS1
CDKN2C
CDR2
CDR2L
CDR3
CDRT1
CDRT15
CDRT15L2
CDRT7
CDV3
CDYL
CEACAM16
CEACAM3
CEACAM4
CEBPB
CEBPG
CECR7
CELF4
CELF5
CELP
CELSR1
CELSR2
CEMIP
CENPBD1P1
CENPL
CEP128
CEP350
CEP44
CEP68
CEP70
CEP85
CEP89
CERS6
CES1P1
CES1P2
CES3
CES4A
CES5AP1
CFDP1
CFLAR
CFLAR-AS1
CFTR
CGB1
CGNL1
CHAC1
CHAD
CHAF1B
CHCHD5
CHCHD6
CHD2
CHD5
CHD6
CHD9
CHEK2P2
CHFR
CHIC2
CHKA
CHL1
CHMP2B
CHN1
CHN2
CHORDC1
CHP1
CHRDL2
CHRM3
CHRM5
CHRNB3
CHST1
CHST11
CHST15
CHST2
CHST3
CHST7
CHUK
CHURC1
CHURC1-FNTB
CIART
CIDEC
CIDECP
CIITA
CILP
CIRBP-AS1
CIRH1A
CISH
CISTR
CIT
CITED2
CITED4
CKB
CKMT2
CKMT2-AS1
CLASP1
CLASP2
CLCC1
CLDN10
CLDN10-AS1
CLDN11
CLDN14
CLDN23
CLDN24
CLDN3
CLDN7
CLDND1
CLEC10A
CLEC16A
CLEC18A
CLEC18C
CLIC3
CLIC4
CLIP1
CLMN
CLMP
CLN6
CLPSL1
CLPSL2
CLPTM1L
CLPX
CLSTN3
CLTCL1
CLU
CLUH
CLVS2
CLYBL
CMBL
CMIP
CMKLR1
CMSS1
CMTM8
CNBP
CNEP1R1
CNGA3
CNGA4
CNGB1
CNN2
CNNM1
CNNM2
CNOT4
CNPY1
CNPY3
CNR2
CNST
CNTD1
CNTN1
CNTN5
CNTNAP3
CNTRL
COA1
COBL
COBLL1
COG1
COG7
COL13A1
COL15A1
COL16A1
COL18A1
COL18A1-AS1
COL1A1
COL1A2
COL21A1
COL22A1
COL23A1
COL26A1
COL27A1
COL4A1
COL4A2
COL4A2-AS1
COL4A3
COL4A4
COL4A6
COL5A1
COL5A2
COL6A2
COL6A3
COLCA2
COLEC12
COMMD4
COMMD6
COMP
COMTD1
COPS3
COPS8
COPZ2
COQ2
CORO1C
CORO2A
CORO2B
CORO6
CORO7
CORO7-PAM16
CORT
COTL1
COX10
COX10-AS1
COX7A2L
CP
CPAMD8
CPB2
CPB2-AS1
CPD
CPEB3
CPEB4
CPLX2
CPM
CPNE4
CPNE7
CPNE8
CPO
CPS1
CPSF2
CPT1A
CPT2
CRADD
CRB1
CRCP
CRCT1
CREB3L1
CREB3L2
CREB3L3
CREB5
CREBBP
CREG1
CRELD2
CRIM1
CRIPT
CRISPLD2
CRK
CRKL
CRLF2
CRMP1
CROCC
CRTAC1
CRTAM
CRTC1
CRY1
CRY2
CRYAA
CRYAB
CRYBA2
CRYBB2P1
CRYL1
CRYM
CRYM-AS1
CRYZ
CSAD
CSF1
CSF3
CSF3R
CSGALNACT1
CSMD2
CSNK1A1
CSNK1D
CSPG4
CSRNP1
CST1
CST3
CSTF3-AS1
CT62
CTAGE1
CTAGE10P
CTAGE5
CTB-174D11.1
CTBP1
CTBP1-AS
CTBP2
CTC-436P18.1
CTCF
CTD-2127H9.1
CTD-2194D22.4
CTD-2258A20.5
CTD-3080P12.3
CTDP1
CTDSP1
CTDSPL
CTF1
CTGF
CTGLF12P
CTIF
CTNNA1
CTNNA3
CTNNB1
CTNNBL1
CTNND1
CTPS1
CTPS2
CTRB2
CTSB
CTSC
CTSH
CTTN
CUEDC1
CUL2
CUX1
CUX2
CWC22
CXADR
CXCL13
CXCL2
CXCL6
CXXC1
CXXC4
CXorf27
CXorf56
CYB561
CYB5A
CYB5B
CYB5R3
CYFIP2
CYGB
CYP1B1
CYP1B1-AS1
CYP21A2
CYP24A1
CYP27A1
CYP2A13
CYP2A6
CYP2A7
CYP2A7PC
CYP2B6
CYP2C18
CYP2C19
CYP2C9
CYP2F1
CYP2G1P
CYP3A4
CYP3A43
CYP3A5
CYP3A7
CYP3A7-CYP3AP1
CYP4B1
CYP4F12
CYP4F22
CYP4F3
CYR61
CYSTM1
CYTH1
CYTH3
CYYR1
D2HGDH
DAAM1
DAAM2
DAB2
DAG1
DAGLA
DALRD3
DAP
DAPK1
DAPK2
DARS2
DAW1
DBP
DCAF12
DCAF12L1
DCAF17
DCAF6
DCAKD
DCBLD1
DCBLD2
DCC
DCDC1
DCDC2
DCDC2C
DCDC5
DCLK2
DCP1A
DCST1
DCST2
DCUN1D1
DCUN1D3
DDAH1
DDC
DDHD1
DDIT4
DDR1
DDT
DDX11
DDX11L9
DDX12P
DDX19A
DDX19B
DDX3X
DDX60
DDX60L
DEAF1

DEC1
DECR1
DEK
DENND1A
DENND2A
DENND3
DENND5B
DENND6B
DEPDC5
DEPTOR
DERA
DFNA5
DFNB31
DGCR8
DGKD
DGKH
DGUOK
DGUOK-AS1
DHCR24
DHCR7
DHODH
DHRS12
DHRS3
DHRS4
DHRS4L1
DHRS4L2
DHRS7B
DHRS9
DHRSX
DHX15
DHX32
DHX36
DHX40
DHX8
DIAPH1
DIAPH2
DIDO1
DIO2
DIO2-AS1
DIO3AS
DIO3OS
DIP2B
DIP2C
DIRC3
DISC1
DISP1
DIXDC1
DKK1
DKK3
DLC1
DLEU1
DLG2
DLG3
DLG5
DLGAP1
DLGAP1-AS1
DLGAP1-AS4
DLK1
DLL4
DLX4
DM119512
DMBX1
DMC1
DNAH11
DNAH12
DNAH14
DNAH17
DNAH5
DNAH7
DNAH9
DNAI2
DNAJB1
DNAJB2
DNAJB5
DNAJB6
DNAJC1
DNAJC12
DNAJC15
DNAJC16
DNAJC5
DNAJC5B
DNAL4
DNASE2B
DNER
DNM1
DNM1P34
DNM1P46
DNM2
DNM3
DNMBP
DNMBP-AS1
DOCK1
DOCK10
DOCK11
DOCK2
DOCK4
DOCK5
DOCK6
DOCK8
DOCK9
DOCK9-AS2
DOK4
DOK6
DONSON
DOT1L
DPEP1
DPF1
DPH5
DPH6
DPH6-AS1
DPM2
DPP10
DPP3
DPP4
DPPA4
DPT
DPY19L1
DPY19L1P1
DPY30
DPYD
DPYD-AS1
DPYSL2
DPYSL3
DRAM1
DRC1
DRGX
DROSHA
DSCAML1
DSCR4
DSCR8
DSG2
DSP
DST
DSTN
DTD2
DTNA
DTNB
DTWD1
DTX2
DTX2P1-UPK3BP1-PMS2P11
DUS3L
DUSP1
DUSP10
DUSP14
DUSP16
DUSP26
DUSP27
DUSP4
DUSP5
DUSP6
DUSP7
DUX4
DUX4L4
DVL2
DYM
DYNC1I1
DYNC2H1
DYNLL1
DYNLT3
DYRK1A
DYRK1B
DYSF
DYX1C1
DZIP1
DZIP1L
E11280
E2F7
EAPP
EBF1
EBI3
EBNA1BP2
ECE1
ECHDC3
ECI1
EDAR
EDARADD
EDC3
EDEM2
EDN1
EDN2
EDN3
EEA1
EEF1E1
EEF1G
EEF2K
EEFSEC
EEPD1
EFCAB1
EFCAB11
EFCAB12
EFCAB13
EFCAB2
EFCAB4A
EFCAB4B
EFCAB6
EFEMP2
EFHB
EFHC1
EFHD1
EFHD2
EFNA1
EFNA3
EFNA5
EFR3A
EFR3B
EGF
EGFL6
EGFLAM
EGFR
EGLN1
EGLN3
EHBP1
EHD1
EHD2
EHD4
EHF
EHHADH-AS1
EHMT1
EHMT1-IT1
EIF1B-AS1
EIF2AK3
EIF2S1
EIF2S3
EIF3A
EIF3H
EIF3M
EIF4E1B
EIF4EBP1
EIF4ENIF1
EIF4G1
EIF4G3
ELANE
ELAVL4
ELF1
ELK3
ELL
ELL2
ELMO1
ELMSAN1
ELN
ELOVL5
ELOVL6
ELP5
EMB
EMC3
EMC3-AS1
EML4
EML6
EMP1
EMP2
EMX2
EMX2OS
EN2
ENAH
ENKUR
ENO1
ENO1-AS1
ENOX1
ENPEP
ENPP6
ENSA
ENTPD1
ENTPD2
ENTPD4
ENTPD6
EP400
EPAS1
EPB41
EPB41L1
EPB41L2
EPB41L4A
EPB41L4B
EPB41L5
EPC1
EPC2
EPDR1
EPHA2
EPHA5
EPHB2
EPHB3
EPHB4
EPM2A
EPN2
EPS15L1
EPS8
EPSTI1
ERBB2IP
ERBB4
ERC1
ERCC1
ERCC4
ERCC6
ERCC6-PGBD3
EREG
ERF
ERGIC1
ERI2
ERI3
ERI3-IT1
ERICH1
ERN1
ERRFI1
ESR1
ESRRG
ESYT2
ETAA1
ETFB
ETFDH
ETNK2
ETS1
ETS2
ETV3
ETV4
ETV6
EVA1A
EVA1C
EVI5
EVL
EVX1
EXOC4
EXOC6
EXOC6B
EXT1
EXT2
EXTL1
EYA2
EYA4
EYS
EZR
F11-AS1
F11R
F2
F2R
F2RL1
F2RL3
F3
F5
F8
FA2H
FABP2
FADD
FADS2
FAIM2
FAM101A
FAM102B
FAM104A
FAM104B
FAM105A
FAM107A
FAM107B
FAM115C
FAM117A
FAM120C
FAM129A
FAM129B
FAM131C
FAM133B
FAM135A
FAM138A
FAM13A
FAM151A
FAM153A
FAM153B
FAM153C
FAM156A
FAM156B
FAM157A
FAM159A
FAM160B1
FAM161A
FAM167A
FAM168A
FAM173B
FAM174B
FAM177A1
FAM180A
FAM182B
FAM183B
FAM183CP
FAM185A
FAM186A
FAM196B
FAM19A2
FAM19A5
FAM207A
FAM209B
FAM20C
FAM210B
FAM214A
FAM220A
FAM222B
FAM223A
FAM227A
FAM227B
FAM228B
FAM230B
FAM25A
FAM25G
FAM27L
FAM35A
FAM35BP
FAM35DP
FAM41C
FAM46A
FAM46B
FAM47B
FAM49A
FAM49B
FAM50B
FAM53B
FAM55B
FAM58A
FAM65B
FAM65C
FAM71A
FAM73A
FAM81A
FAM83A
FAM86C2P
FAM86DP
FAM86FP
FAM86HP
FAM89A
FAM96A
FAM9B
FAM9C
FANCC
FANCD2
FARP1
FARS2
FASTKD2
FASTKD5
FAT1
FAT3
FAXDC2
FBLIM1
FBLN1
FBLN5
FBN1
FBN2
FBP1
FBRSL1
FBXL13
FBXL14
FBXL16
FBXL17
FBXL18
FBXL21
FBXO15
FBXO16
FBXO18
FBXO21
FBXO32
FBXO34
FBXO36
FBXO42
FBXO46
FBXO7
FBXO8
FBXW10
FBXW2
FBXW5
FBXW7
FCER1A
FCGBP
FCHO1
FCHO2
FCHSD2
FCN2
FDXACB1
FEM1C
FER1L5
FER1L6
FER1L6-AS1
FER1L6-AS2
FERMT2
FGA
FGB
FGD2
FGD4
FGD5
FGD6
FGF1
FGF18
FGF2
FGF20
FGF7
FGFBP3
FGFR2
FGGY
FGL1
FHAD1
FHDC1
FHIT
FHL1
FIBCD1
FIG4
FIGF
FIGN
FIGNL2
FILIP1
FILIP1L
FITM2

FKBP11
FKBP5
FKBP6
FLNB
FLRT2
FLT3
FLVCR1
FLVCR2
FMN1
FMN2
FMNL2
FMOD
FN1
FNBP1
FNBP1L
FNDC3A
FNDC3B
FNDC8
FNIP1
FNIP2
FNTB
FOCAD
FOLR1
FOPNL
FOSL2
FOXD2
FOXD4L5
FOXE1
FOXI1
FOXJ1
FOXJ3
FOXK1
FOXK2
FOXM1
FOXN2
FOXN3
FOXN4
FOXO1
FOXO3
FOXP1
FOXP2
FOXP4
FOXQ1
FOXS1
FP15737
FRAS1
FRAT2
FRG2
FRK
FRMD1
FRMD3
FRMD4A
FRMD6
FRMD6-AS2
FRMPD4
FRS2
FRYL
FSCN1
FSIP1
FSTL3
FSTL4
FSTL5
FTO
FTX
FUCA2
FUNDC2
FURIN
FUT3
FUT5
FXN
FXYD4
FYB
FZD2
FZD5
FZD7
G6PC
G6PD
GAA
GAB2
GAB3
GABARAPL1
GABRB3
GABRE
GABRG3
GADD45A
GADD45B
GADD45G
GADL1
GAL
GALK1
GALM
GALNS
GALNT1
GALNT10
GALNT13
GALNT15
GALNT16
GALNT18
GALNT2
GALNT5
GALNT9
GALNTL5
GALNTL6
GALP
GALR1
GANC
GAPDHS
GAREM
GARNL3
GAS2L1
GAS6
GAS6-AS1
GAS7
GATA3
GATA6
GATAD2B
GATS
GATSL2
GBE1
GBF1
GBGT1
GCC2
GCH1
GCM1
GCN1L1
GCOM1
GCSH
GDAP1L1
GDAP2
GDF10
GDF15
GDF5
GDPD1
GDPD5
GEM
GEMIN7
GEMIN8
GFM1
GFOD1
GFPT1
GFPT2
GFRA1
GFRA2
GGA2
GGH
GGT1
GGT3P
GGT5
GGTLC1
GGTLC2
GH1
GHRHR
GINS3
GIP
GIPC1
GIT1
GJA1
GJB3
GJC1
GJD2
GJD3
GKN1
GLB1
GLCE
GLDC
GLG1
GLI2
GLIS3
GLP2R
GLRX
GLS
GLTP
GLTSCR1
GMNC
GMPR
GNA12
GNA13
GNA15
GNAI1
GNAI3
GNAL
GNAO1
GNAQ
GNAT3
GNB1
GNG12
GNG12-AS1
GNG7
GNLY
GNMT
GNN
GOLGA4
GOLGA6A
GOLGA6L2
GOLGA6L4
GOLGA6L6
GOLGA8CP
GOLGA8DP
GOLGA8I
GOLGA8M
GOLGA8S
GOLGA8T
GOLM1
GOPC
GOSR2
GOT1
GPAM
GPATCH1
GPATCH8
GPC1
GPC6
GPC6-AS2
GPCPD1
GPD2
GPIHBP1
GPNMB
GPR1
GPR1-AS
GPR108
GPR110
GPR111
GPR115
GPR116
GPR123
GPR126
GPR132
GPR133
GPR148
GPR153
GPR158
GPR173
GPR176
GPR39
GPR4
GPR56
GPR64
GPR68
GPR75
GPR75-ASB3
GPR87
GPR98
GPRC5A
GPRIN2
GPRIN3
GPSM1
GPT2
GRAMD1A
GRAMD1B
GRAMD2
GRAMD3
GRAMD4
GRAP
GRAP2
GRAPL
GRASP
GRB10
GRB2
GREB1
GREB1L
GRHL2
GRHL3
GRHPR
GRID1
GRID2
GRIK4
GRIN2B
GRIP1
GRK5
GRM5
GRTP1
GS1-122H1.2
GS1-600G8.3
GSAP
GSDMC
GSDMD
GSE1
GSG1
GSN
GSN-AS1
GSR
GSTA3
GSTA7P
GSTTP2
GTDC1
GTF2F2
GTF2I
GTF2IP1
GTF2IRD1
GTF2IRD2
GTF3C5
GTF3C6
GUCA1A
GUCA2A
GUCD1
GUCY2EP
GUCY2GP
GULP1
GUSBP1
GUSBP11
GUSBP2
GUSBP9
GXYLT2
GYG1
GYG2
GYS2
GZMM
H3F3C
HACE1
HAL
HAND1
HAPLN2
HAS2
HAVCR1
HAVCR1P1
HAVCR2
HBE1
HBG2
HCAR1
HDAC11
HDAC4
HDAC7
HDAC9
HDLBP
HEATR2
HEATR4
HEATR5A
HEBP2
HECA
HECTD1
HECTD2
HECTD4
HECW1
HECW2
HELZ
HELZ2
HEMK1
HERC1
HERC2
HERC2P2
HERC2P3
HERC2P7
HERC2P9
HERC3
HERC4
HERPUD1
HES1
HES5
HEXDC
HFM1
HGD
HGF
HGSNAT
HHAT
HHEX
HIC2
HID1
HIF3A
HILPDA
HIP1
HIPK2
HIST1H1D
HIST1H2BD
HIVEP1
HIVEP2
HIVEP3
HK1
HK2
HLA-G
HLA-J
HLCS
HLF
HM13
HM13-AS1
HM222546
HMBOX1
HMCN1
HMGA1P7
HMGA2
HMGB3
HMGCL
HMGCS1
HMHA1
HMHB1
HMOX1
HMOX2
HN1L
HNF1A
HNF1B
HNF4A
HNF4A-AS1
HNF4G
HNMT
HNRNPD
HNRNPH2
HOGA1
HOMER1
HOMER2
HOOK2
HOOK3
HORMAD2
HOXB-AS1
HOXB-AS3
HOXB1
HOXB3
HOXB6
HOXC13
HP11097
HPCAL1
HPD
HPGD
HPGDS
HPN
HPRT1
HPS3
HPSE2
HPVC1
HRASLS2
HRCT1
HRG
HS1BP3
HS1BP3-IT1
HS3ST1
HS3ST3B1
HSD11B1
HSD11B2
HSD17B3
HSD17B4
HSD17B6
HSD3BP4
HSD52
HSF2BP
HSFY1P1
HSPB3
HSPB7
HSPB8
HSPG2
HSPH1
HTATIP2
HTR1B
HTR1D
HTR3E
HTRA3
HTT
HUWE1
IAH1
IARS
ICA1L
ICOSLG
ID1
ID3
IDH1
IDI2-AS1
IDO2
IDUA
IER3
IER5
IFFO2
IFI35
IFIT1
IFITM10
IFLTD1
IFNAR1
IFNAR2
IFNG-AS1
IFRD1
IFT122
IFT140
IFT80
IFT81
IGF1
IGF1R
IGF2BP1
IGF2BP2
IGF2BP3
IGFBP1
IGFBP2
IGFBP3
IGFBP4
IGFBP5
IGFBP7
IGHD
IGHE
IGHG1
IGHV1-18
IGLL1
IGSF1
IGSF11
IGSF21
IGSF23
IGSF3
IGSF9B
IKBKG
IL13RA1
IL15
IL17RC
IL17REL
IL18
IL18RAP
IL1A
IL1R1
IL1R2
IL1RAP
IL1RL2
IL21-AS1

IL22RA1
IL24
IL3RA
IL6R
IL6ST
IL8
IMMP2L
IMP3
IMPA2
IMPDH1
IMPG2
INHBB
INIP
INMT
INO80C
INPP1
INPP4A
INPP4B
INPP5A
INPP5B
INPP5D
INPP5J
INSL4
INSR
INTS1
INTS3
INTS4
INTS7
INTS9
INVS
IP6K2
IP6K3
IPMK
IPO11
IPO11-LRRC70
IPO5
IPO7
IPO8
IPO9-AS1
IQCB1
IQCD
IQCH
IQCJ
IQCJ-SCHIP1
IQCK
IQGAP1
IQGAP2
IQSEC1
IQSEC2
IRAK2
IRAK3
IRF2BP2
IRF4
IRF8
IRS1
IRS2
IRX3
ISCA1
ISG20
ISL1
ITCH
ITGA10
ITGA11
ITGA2
ITGA3
ITGA5
ITGA6
ITGA7
ITGA8
ITGA9
ITGAE
ITGB1
ITGB2
ITGB2-AS1
ITGB3
ITGB4
ITGB5
ITGB6
ITGB8
ITGBL1
ITIH2
ITLN2
ITM2C
ITPA
ITPK1
ITPKA
ITPKB
ITPKC
ITPR1
ITPR2
ITPR3
ITPRIP
ITPRIPL2
ITSN1
ITSN2
IWS1
IYD
IZUMO2
JADE1
JADE2
JAG1
JAK1
JARID2
JAZF1
JDP2
JHDM1D-AS1
JOSD2
JPH2
JPH4
JPX
JRK
JUNB
JUP
KALRN
KANK1
KANK4
KANSL1
KANSL1L
KAT2B
KAT6A
KAT6B
KAT8
KATNAL1
KAZN
KBTBD2
KCMF1
KCNAB1
KCNB1
KCNC2
KCNE3
KCNE4
KCNG1
KCNH2
KCNH5
KCNH7
KCNIP1
KCNIP3
KCNJ12
KCNJ15
KCNJ2
KCNJ4
KCNJ6
KCNK1
KCNK17
KCNK3
KCNK5
KCNMA1
KCNN3
KCNQ1
KCNQ1DN
KCNQ5
KCNS2
KCNU1
KCTD1
KCTD16
KCTD17
KCTD3
KCTD9
KDELR3
KDM3A
KDM4B
KDM4C
KDM6B
KDR
KIAA0020
KIAA0226
KIAA0232
KIAA0355
KIAA0368
KIAA0391
KIAA0430
KIAA0513
KIAA0556
KIAA0753
KIAA0825
KIAA0922
KIAA0930
KIAA1033
KIAA1107
KIAA1147
KIAA1210
KIAA1211
KIAA1211L
KIAA1217
KIAA1279
KIAA1324
KIAA1377
KIAA1432
KIAA1462
KIAA1522
KIAA1549
KIAA1671
KIAA1683
KIAA1755
KIAA1919
KIDINS220
KIF12
KIF13A
KIF13B
KIF16B
KIF17
KIF19
KIF1B
KIF21B
KIF24
KIF26A
KIF2B
KIF3C
KIF5C
KIFC3
KIR2DL4
KIR2DS2
KIR2DS4
KIR2DS5
KIR3DL1
KIR3DL2
KIR3DS1
KIT
KITLG
KL
KLC2
KLF13
KLF15
KLF2
KLF3
KLF3-AS1
KLF4
KLF5
KLF6
KLF7
KLF9
KLHDC4
KLHL13
KLHL2
KLHL25
KLHL26
KLHL29
KLHL30
KLHL35
KLHL36
KLHL38
KLHL4
KLK13
KLRC1
KLRC2
KLRC3
KLRC4
KLRC4-KLRK1
KLRK1
KMT2C
KPNA2
KPNA3
KPNA4
KPNA7
KPNB1
KRBA1
KRR1
KRT12
KRT121P
KRT14
KRT16
KRT17
KRT18
KRT19P2
KRT20
KRT28
KRT3
KRT34
KRT37
KRT39
KRT4
KRT42P
KRT5
KRT6A
KRT7
KRT78
KRT8
KRT80
KRT86
KRTAP10-12
KRTAP10-3
KRTAP17-1
KRTAP2-3
KRTAP21-2
KRTAP4-11
KRTAP4-5
KRTAP4-7
KRTAP5-11
KRTAP9-7
KSR1
KSR2
KTN1
KTN1-AS1
KU-MEL-3
KY
KYNU
L1TD1
LACTB2
LAD1
LAMA2
LAMA3
LAMA4
LAMA5
LAMB1
LAMB4
LAMC1
LAMC2
LAMC3
LAMTOR5-AS1
LAPTM4A
LAPTM5
LARGE
LARP1
LARP4B
LARS2
LATS2
LBH
LBP
LBR
LCA5
LCE5A
LCN12
LCN9
LCNL1
LCOR
LCP1
LCP2
LDHA
LDHAL6A
LDHB
LDLR
LDLRAD3
LDLRAD4
LDLRAP1
LECT2
LEKR1
LEMD1-AS1
LEMD2
LEO1
LEP
LEPR
LEPREL1
LEPREL4
LEPROT
LEPROTL1
LFNG
LGALS3
LGALS4
LGALSL
LGI4
LGMN
LGR4
LGR5
LGR6
LHFPL2
LHFPL5
LHPP
LHX4
LIF
LIFR-AS1
LILRA6
LILRB1
LILRB3
LILRB4
LIMA1
LIMCH1
LIMD1
LIMK1
LIMS1
LIN28B
LINC-PINT
LINC00111
LINC00113
LINC00114
LINC00152
LINC00160
LINC00189
LINC00211
LINC00229
LINC00239
LINC00240
LINC00251
LINC00263
LINC00273
LINC00277
LINC00282
LINC00284
LINC00299
LINC00313
LINC00316
LINC00319
LINC00322
LINC00324
LINC00326
LINC00330
LINC00332
LINC00336
LINC00351
LINC00417
LINC00426
LINC00457
LINC00469
LINC00470
LINC00473
LINC00474
LINC00476
LINC00479
LINC00486
LINC00499
LINC00501
LINC00502
LINC00511
LINC00520
LINC00523
LINC00525
LINC00534
LINC00548
LINC00570
LINC00572
LINC00578
LINC00589
LINC00592
LINC00595
LINC00598
LINC00602
LINC00605
LINC00607
LINC00615
LINC00628
LINC00637
LINC00640
LINC00642
LINC00664
LINC00670
LINC00671
LINC00673
LINC00674
LINC00675
LINC00676
LINC00689
LINC00690
LINC00702
LINC00703
LINC00704
LINC00706
LINC00708
LINC00837
LINC00838
LINC00842
LINC00844
LINC00853
LINC00856
LINC00862
LINC00866
LINC00869
LINC00875
LINC00877
LINC00879
LINC00880
LINC00886
LINC00887
LINC00894
LINC00895
LINC00911
LINC00920
LINC00926
LINC00928
LINC00933
LINC00937
LINC00940
LINC00941
LINC00942
LINC00950
LINC00954
LINC00961
LINC00963
LINC00970
LINC00999
LINC01006
LINC01010
LINC01030
LINC01057
LINC01080
LINC01085
LINC01093
LINC01094
LINC01101
LINC01108
LINC01119
LINC01121
LINC01122
LINC01123
LINC01132
LINC01135
LINC01137
LINC01138
LINC01146
LINC01159
LINC01169
LINC01184
LINC01185
LINC01186
LINC01191
LINC01197

LINC01203
LINC01212
LINC01213
LINC01214
LINC01220
LINC01234
LINC01237
LINC01252
LINC01257
LINC01260
LINGO2
LINGO3
LIPA
LIPC
LIPE-AS1
LITAF
LIX1
LLGL2
LMAN2
LMBR1
LMCD1-AS1
LMF1
LMNA
LMO3
LMO7
LMOD1
LMTK2
LMTK3
LNX1
LNX1-AS2
LOC100093631
LOC100128233
LOC100129316
LOC100130075
LOC100130539
LOC100130872
LOC100131138
LOC100131347
LOC100131496
LOC100132354
LOC100132735
LOC100132891
LOC100133050
LOC100133091
LOC100134868
LOC100188947
LOC100287015
LOC100287072
LOC100287534
LOC100287944
LOC100288255
LOC100288637
LOC100288974
LOC100499194
LOC100505841
LOC100505984
LOC100506022
LOC100506083
LOC100506470
LOC100506714
LOC100506746
LOC100506860
LOC100507156
LOC100507217
LOC100507351
LOC100507424
LOC100507468
LOC100616530
LOC100652824
LOC100862671
LOC100996255
LOC100996385
LOC101926889
LOC101926935
LOC101926940
LOC101926966
LOC101927181
LOC101927526
LOC101927722
LOC101927934
LOC101928136
LOC101928154
LOC101928269
LOC101928283
LOC101928401
LOC101928569
LOC101928600
LOC101928618
LOC101929082
LOC101929125
LOC101929154
LOC101929625
LOC101929681
LOC102031319
LOC102467147
LOC102467213
LOC102467226
LOC102477328
LOC102546229
LOC145783
LOC146880
LOC148696
LOC151760
LOC152586
LOC154761
LOC158434
LOC257396
LOC283685
LOC283767
LOC284412
LOC284865
LOC285593
LOC285768
LOC286190
LOC341056
LOC344967
LOC388553
LOC388882
LOC388906
LOC399715
LOC400548
LOC400655
LOC400891
LOC401177
LOC401320
LOC401557
LOC440040
LOC440243
LOC440311
LOC440434
LOC440461
LOC441242
LOC442132
LOC541472
LOC553103
LOC642776
LOC643441
LOC643486
LOC643770
LOC644172
LOC645513
LOC645752
LOC646626
LOC646736
LOC646743
LOC648987
LOC650368
LOC653501
LOC728323
LOC728819
LOC729080
LOC729444
LOC729737
LOC730102
LOC730668
LOC731779
LOC90784
LOH12CR1
LONRF2
LOXL1
LOXL2
LPA
LPAR1
LPCAT1
LPGAT1
LPHN1
LPHN3
LPIN1
LPIN2
LPIN3
LPP
LPP-AS2
LPPR1
LRBA
LRCH1
LRFN2
LRG1
LRGUK
LRIG1
LRP11
LRP1B
LRP3
LRP5
LRP6
LRP8
LRPPRC
LRRC1
LRRC16A
LRRC25
LRRC28
LRRC29
LRRC36
LRRC37A3
LRRC37A5P
LRRC37B
LRRC37BP1
LRRC48
LRRC49
LRRC56
LRRC59
LRRC74
LRRC8A
LRRC8D
LRRFIP2
LRRIQ1
LRRK2
LRRc37A
LSAMP
LSM1
LSM6
LSP1
LSS
LTBP1
LTBP2
LTBR
LURAP1L
LUZP6
LY6K
LYN
LYPLAL1
LYRM1
LYRM4
LYRM9
LYSMD3
LYZL1
LYZL2
LZTFL1
M1AP
MAB21L3
MACROD2
MAD1L1
MAEL
MAFA
MAFF
MAFK
MAG
MAGEA4
MAGI1
MAGI2
MAGI2-AS3
MAL
MALAT1
MALL
MALSU1
MAML2
MAML3
MAN1A1
MAN1C1
MANBA
MAOA
MAP1B
MAP1LC3B2
MAP2
MAP2K1
MAP2K2
MAP2K4P1
MAP2K6
MAP3K1
MAP3K14
MAP3K14-AS1
MAP3K5
MAP3K6
MAP3K7
MAP3K7CL
MAP3K8
MAP3K9
MAP4
MAP4K3
MAP4K4
MAP7
MAP7D1
MAP7D2
MAPK1
MAPK10
MAPK4
MAPK6
MAPKAP1
MAPKAPK2
MAPRE1
MAPRE2
MAPRE3
MARCH10
MARCH2
MARCH5
MARCH8
MARK2
MARS
MARVELD1
MAS1
MAST2
MAST4
MAT1A
MAU2
MAX
MB21D2
MBD2
MBNL1
MBNL1-AS1
MBNL2
MBOAT2
MBP
MC5R
MCF2L2
MCFD2
MCHR1
MCM5
MCM8
MDFIC
MDGA1
MDH2
MDM1
MDS2
ME3
MEAF6
MECP2
MED12L
MED13
MED13L
MED15
MED16
MED18
MED24
MED26
MED27
MED29
MED30
MED9
MEF2A
MEF2D
MEG9
MEGF9
MEMO1
MEOX1
MERTK
MET
METRNL
METTL15
METTL20
METTL4
METTL9
MFAP2
MFAP4
MFGE8
MFI2
MFNG
MFSD2A
MFSD2B
MFSD6
MFSD6L
MGAM
MGAT3
MGAT4A
MGAT5B
MGLL
MGME1
MGMT
MGRN1
MGST1
MGST3
MICA
MICAL1
MICAL2
MICAL3
MICALL1
MICALL2
MID1
MIG7
MINA
MINOS1
MINOS1-NBL1
MIPEP
MIPEPP3
MIPOL1
MIR100HG
MIR1208
MIR122
MIR1253
MIR1268A
MIR129-1
MIR1302-4
MIR1303
MIR138-2
MIR1538
MIR1587
MIR193B
MIR2054
MIR210HG
MIR2117
MIR222
MIR30A
MIR3141
MIR3156-3
MIR3159
MIR3175
MIR3193
MIR31HG
MIR3201
MIR3649
MIR3685
MIR3686
MIR3688-1
MIR376A1
MIR378D2
MIR3925
MIR3927
MIR3929
MIR3936
MIR3978
MIR4276
MIR4289
MIR4302
MIR4303
MIR4309
MIR4418
MIR4425
MIR4430
MIR4431
MIR4436A
MIR4454
MIR4462
MIR4464
MIR4465
MIR4468
MIR4472-2
MIR4478
MIR4494
MIR450B
MIR4521
MIR4522
MIR4530
MIR4532
MIR4645
MIR4660
MIR4663
MIR4673
MIR4675
MIR4686
MIR4703
MIR4708
MIR4720
MIR4737
MIR4757
MIR4795
MIR4799
MIR5091
MIR5093
MIR5096
MIR5188
MIR5197
MIR548AI
MIR548AJ1
MIR548AP
MIR548F3
MIR548N
MIR5580
MIR5684
MIR5689
MIR5700
MIR5703
MIR5705
MIR583
MIR589
MIR6070
MIR6078
MIR6087
MIR6089-1
MIR6131
MIR633
MIR649
MIR670
MIR6732
MIR6859-1
MIR6860
MIR7641-2
MIR7848
MIR8052
MIR8068
MIR8070
MIR8072
MIR8079
MIR941-1
MIRLET7BHG
MIRLET7I
MISP
MITF
MKKS
MKL1
MKL2
MKLN1
MKNK1
MKNK2
MKRN9P
MLEC
MLH3
MLIP
MLK7-AS1
MLLT3
MLLT4
MLLT6
MLNR
MLPH
MLXIP
MLXIPL
MMAA

MMD
MMD2
MMP15
MMP17
MMP19
MMP20
MMP24
MMP24-AS1
MMP25
MMP28
MMP7
MNAT1
MND1
MOB2
MOB3A
MOB3B
MOB3C
MOCS1
MOGAT1
MOGAT3
MOK
MON2
MORC4
MORF4L2
MOV10
MPC2
MPDZ
MPRIP
MPST
MPV17
MPV17L
MPZL1
MR1
MRDS1
MRE11A
MRFAP1
MRO
MROH1
MROH2A
MROH5
MRPL1
MRPL15
MRPL34
MRPL44
MRPL45P2
MRPL46
MRPL57
MRPL9
MRPS18A
MRPS2
MRPS22
MRPS23
MRPS28
MRPS35
MRPS6
MRS2
MRVI1
MSI1
MSI2
MSL3
MSLN
MSN
MSRA
MSRB1
MST1P2
MSX2
MT1A
MT1DP
MT1X
MT2A
MT4
MTA2
MTAP
MTCL1
MTDH
MTF2
MTFP1
MTHFD1L
MTHFD2L
MTHFS
MTIF3
MTMR10
MTMR12
MTMR2
MTNR1A
MTNR1B
MTOR
MTPN
MTR
MTRNR2L1
MTRNR2L7
MTRNR2L8
MTSS1
MTURN
MTUS1
MTX1
MUC1
MUC12
MUC16
MUC17
MUC20
MUC5B
MUCL1
MUL1
MUSK
MVB12A
MXD1
MXD4
MYADM
MYBL1
MYBPC1
MYBPH
MYC
MYCBP2
MYCN
MYEOV
MYEOV2
MYH10
MYH14
MYH3
MYH9
MYL2
MYL3
MYLK
MYNN
MYO10
MYO15A
MYO16
MYO18B
MYO1A
MYO1B
MYO1D
MYO1E
MYO1G
MYO3B
MYO5A
MYO5C
MYO7B
MYO9B
MYOC
MYOF
MYOG
MYOM1
MYOM2
MYOM3
MYOT
MYOZ2
MYPN
MYRIP
MYT1
MYT1L
MYZAP
N4BP2
NAA20
NAALAD2
NAALADL2
NAB1
NABP1
NACA
NACA2
NACC2
NAIF1
NAIP
NAMA
NAMPT
NANOG
NANOGP1
NAP1L1
NARFL
NARG2
NAT16
NAV1
NAV2
NAV3
NBAS
NBEAL2
NBEAP1
NBL1
NBN
NBPF1
NBPF10
NBPF11
NBPF12
NBPF13P
NBPF14
NBPF18P
NCALD
NCAM1
NCCRP1
NCF1
NCF1C
NCK2
NCKAP1
NCKAP5
NCMAP
NCOA2
NCOA3
NCOA7
NCOR1
NCOR2
NCS1
NDE1
NDEL1
NDFIP1
NDFIP2
NDRG1
NDRG3
NDRG4
NDST1
NDUFA1
NDUFA10
NDUFA11
NDUFAF6
NDUFV2
NEAT1
NEB
NEBL
NEBL-AS1
NEDD4
NEDD4L
NEDD9
NEFL
NEK10
NEK11
NEK6
NEK7
NELL1
NEMF
NENF
NEO1
NET1
NEU2
NEURL1
NEURL1B
NF1
NFATC1
NFATC2
NFE2L1
NFE2L2
NFIA
NFIB
NFIC
NFKB1
NFKBIA
NFX1
NFYB
NGEF
NHEJ1
NHS
NHSL1
NIFK-AS1
NINJ1
NINJ2
NINL
NIPAL2
NIPBL
NIPSNAP1
NKAIN1
NKAIN3
NKIRAS1
NKPD1
NKX2-6
NLRP2
NME7
NME8
NMNAT2
NMT1
NNMT
NOC3L
NOC4L
NOD1
NOL10
NOL11
NOL3
NOM1
NOS1
NOS1AP
NOS2
NOSTRIN
NOTCH1
NOTCH2
NOTCH2NL
NOTCH3
NOVA2
NOX1
NOX5
NPAS2
NPEPPS
NPFFR1
NPHP4
NPIPA3
NPNT
NPPB
NPR2
NPR3
NPSR1
NPSR1-AS1
NPTX1
NPTX2
NPY4R
NR0B1
NR1D2
NR1I2
NR2C2
NR2F1-AS1
NR2F2
NR2F2-AS1
NR2F6
NR3C1
NR5A1
NR5A2
NRCAM
NRDE2
NREP
NRG1
NRG2
NRIP1
NRP1
NRP2
NRSN1
NRSN2
NRXN3
NS3BP
NSDHL
NSG1
NSMCE1
NSMCE2
NSUN3
NSUN5
NSUN6
NT5C1B
NT5C1B-RDH14
NT5C2
NT5DC4
NTHL1
NTMT1
NTN4
NTRK2
NTRK3
NUAK1
NUAK2
NUB1
NUBP1
NUCKS1
NUDT13
NUDT16
NUDT7
NUDT9P1
NUFIP1
NUP160
NUP188
NUP210
NUP214
NUP62CL
NUP93
NUPR1
NUTM2D
NUTM2F
NWD1
NXPH3
NYAP1
NYAP2
NYX
OACYLP
OAF
OAS1
OAS3
OAZ3
OBFC1
OCA2
OCEL1
ODC1
ODF1
OFCC1
OGDH
OGFRL1
OGT
OLAH
OLFML2A
OLFML3
ONECUT3
OPA3
OPHN1
OPLAH
OPN3
OPN4
OPRL1
OPTC
OR10H5
OR10V1
OR11H1
OR1A2
OR1D2
OR1G1
OR1M1
OR2A1
OR2K2
OR2T1
OR4C46
OR4D1
OR4F16
OR51B5
OR52K1
OR56A5
OR5AK4P
OSBP2
OSBPL10
OSBPL1A
OSBPL3
OSBPL5
OSBPL6
OSBPL8
OSBPL9
OSGIN2
OSM
OSR1
OSTCP1
OTOF
OTOGL
OTOS
OTUD7A
OXR1
OXTR
P2RX1
P2RY1
P2RY2
P2RY8
P4HA2
P4HA3
PABPC1P2
PACRG
PACS1
PACSIN1
PACSIN2
PACSIN3
PADI1
PADI4
PAH
PAK1
PAK4
PAK7
PALLD
PALM
PALM2-AKAP2
PALMD
PANK3
PANX1
PAPD5
PAPL
PAPOLG
PAPPA
PAPSS1
PAPSS2
PAQR5
PAQR6
PAQR8
PAQR9
PARD3
PARD3-AS1
PARD6B
PARGP1
PARK2
PARM1
PARN
PARP11
PARP12
PARP14
PARP15
PARP4
PART1
PARVA
PARVB
PAWR
PAX2
PAX5
PAX7
PAX8
PAX8-AS1
PBRM1
PBX1
PBX4
PC
PCAT14
PCAT18
PCAT19
PCBP1-AS1
PCBP3
PCDH1
PCDH15
PCDH18
PCDH9
PCDHA1
PCDHA10
PCDHA11
PCDHA12
PCDHA13
PCDHA2
PCDHA3
PCDHA4
PCDHA5
PCDHA6
PCDHA7
PCDHA8
PCDHA9

PCDHAC1
PCDHAC2
PCDHB19P
PCDHB2
PCDP1
PCED1B
PCED1B-AS1
PCGF5
PCK1
PCLO
PCNT
PCNX
PCNXL2
PCSK1
PCSK5
PCSK6
PCYT1B
PDC
PDCD6IP
PDDC1
PDE10A
PDE11A
PDE1A
PDE1C
PDE3A
PDE4B
PDE4D
PDE4DIP
PDE5A
PDE6A
PDE7A
PDE7B
PDE8A
PDE8B
PDE9A
PDGFC
PDGFD
PDGFRA
PDGFRB
PDGFRL
PDHX
PDIA5
PDIA6
PDK1
PDK2
PDK4
PDLIM1
PDLIM5
PDP1
PDPK1
PDS5A
PDSS2
PDX1-AS1
PDXDC1
PDXK
PDZD2
PDZD8
PDZK1
PDZRN4
PEAK1
PEAR1
PEBP4
PECAM1
PECR
PELI2
PELP1
PEMT
PER1
PER2
PERP
PEX11A
PEX14
PEX5
PEX5L
PFKFB3
PFKP
PGC
PGF
PGM1
PGM2
PGPEP1
PGRMC2
PGS1
PHACTR2
PHACTR3
PHACTR4
PHB
PHC1
PHC2
PHEX
PHF14
PHF20
PHKA1
PHKA2-AS1
PHKB
PHLDA1
PHLDA3
PHLDB2
PHLPP1
PHLPP2
PHRF1
PHTF2
PHYHD1
PHYHIP
PI4K2A
PIAS1
PIAS4
PICALM
PICK1
PID1
PIEZO2
PIGC
PIGG
PIGL
PIGU
PIK3AP1
PIK3C2B
PIK3C2G
PIK3C3
PIK3CB
PIK3CD
PIK3R1
PIK3R3
PIM1
PIM3
PIN4
PIP4K2A
PIP5K1A
PIP5K1C
PIR-FIGF
PITPNB
PITPNC1
PITPNM2
PITPNM3
PITX3
PIWIL2
PKD1L1
PKD1L2
PKD2L1
PKD2L2
PKDCC
PKIA
PKIG
PKNOX2
PKP1
PKP2
PKP4
PLA2G10
PLA2G16
PLA2G1B
PLA2G4A
PLA2G6
PLA2R1
PLAC1
PLAGL2
PLAT
PLAUR
PLB1
PLCB1
PLCB4
PLCE1
PLCG1
PLCH1
PLCH2
PLCL2
PLCXD1
PLCXD2
PLCXD3
PLD1
PLD5
PLEC
PLEK
PLEKHA2
PLEKHA5
PLEKHA6
PLEKHA7
PLEKHA8P1
PLEKHF2
PLEKHG1
PLEKHG2
PLEKHG3
PLEKHG4B
PLEKHG6
PLEKHH2
PLEKHM1P
PLEKHM3
PLIN2
PLIN3
PLIN5
PLK2
PLK5
PLLP
PLOD1
PLOD2
PLSCR1
PLSCR2
PLSCR3
PLSCR4
PLSCR5
PLVAP
PLXDC1
PLXNA1
PLXNA2
PLXNA3
PLXNA4
PLXNB1
PLXNB2
PLXNC1
PLXND1
PMAIP1
PMEPA1
PMF1
PMF1-BGLAP
PMPCA
PMS2
PMS2L14
PMS2P4
PMS2P5
PMS2P9
PMVK
PNKP
PNLIP
PNMT
PNPLA2
PNPLA3
PNPLA5
PNPLA8
PNPO
POC1B
POC1B-GALNT4
PODXL
POGLUT1
POGZ
POLE2
POLG
POLI
POLR2C
POLR2E
POLR2J3
POLR2M
POLRMT
POM121
POM121L10P
POM121L8P
POMC
POMP
PON1
PON2
PON3
POP5
POR
POTEG
POTEH
POTEM
POU1F1
POU2AF1
PPA1
PPARA
PPARG
PPARGC1A
PPARGC1B
PPCDC
PPEF1
PPEF2
PPFIA1
PPFIBP1
PPFIBP2
PPHLN1
PPIAL4F
PPIH
PPIL2
PPIL4
PPL
PPM1B
PPM1D
PPM1E
PPM1H
PPM1L
PPP1CB
PPP1R12A
PPP1R12B
PPP1R13B
PPP1R14C
PPP1R15B
PPP1R16B
PPP1R21
PPP1R3B
PPP1R3C
PPP1R9A
PPP2CB
PPP2R1B
PPP2R2A
PPP2R2C
PPP2R3C
PPP2R5A
PPP3CA
PPP4R2
PPP5D1
PPP6R3
PPTC7
PRDM1
PRDM11
PRDM15
PRDM16
PRDM2
PRELID2
PREX1
PRICKLE2
PRICKLE2-AS1
PRICKLE2-AS3
PRIMA1
PRKAA2
PRKACB
PRKAG2
PRKAG2-AS1
PRKAR1A
PRKAR1B
PRKCA
PRKCB
PRKCD
PRKCE
PRKCH
PRKCQ
PRKD1
PRKX
PRL
PRLHR
PRMT2
PRMT3
PRMT9
PROC
PRODH
PROP1
PROSER1
PROSER2
PROSER2-AS1
PROX1
PROZ
PRPS1
PRPSAP2
PRR11
PRR12
PRR15
PRR15L
PRR26
PRR5
PRR5-ARHGAP8
PRR5L
PRRC2B
PRRX2
PRSS23
PRSS3
PRUNE2
PRX
PSAP
PSCA
PSD3
PSD4
PSG1
PSG10P
PSG2
PSG3
PSG4
PSG5
PSG6
PSG7
PSG9
PSKH1
PSMA1
PSMA6
PSMB2
PSMB4
PSMB7
PSMD1
PSMD12
PSMD14
PSME4
PSORS1C3
PSTPIP1
PTBP1
PTCHD2
PTDSS2
PTEN
PTGDR
PTGDS
PTGER2
PTGER4
PTGES
PTGIS
PTGS1
PTGS2
PTH1R
PTK2
PTK2B
PTMA
PTP4A1
PTP4A2
PTPDC1
PTPN1
PTPN11
PTPN12
PTPN14
PTPN20A
PTPN20B
PTPN21
PTPN3
PTPN4
PTPN9
PTPRF
PTPRG
PTPRH
PTPRJ
PTPRK
PTPRM
PTPRN2
PTPRO
PTPRQ
PTPRR
PTPRS
PTPRU
PTRH1
PTS
PUM1
PUM2
PURB
PUS7
PVRL1
PVRL4
PVT1
PWRN1
PWWP2B
PXDN
PXMP4
PXN
PXYLP1
PYCR2
PYGB
PYGL
PYROXD2
PYURF
QKI
QRSL1
QSOX1
RAB10
RAB11B
RAB11B-AS1
RAB11FIP1
RAB11FIP2
RAB11FIP3
RAB11FIP4
RAB20
RAB22A
RAB27B
RAB2B
RAB30
RAB30-AS1
RAB31
RAB33A
RAB3B
RAB3D
RAB4A
RAB5C
RAB7A
RABEP1
RABGAP1
RABL2A
RABL3
RAC1
RAC2
RAD18
RAD21
RAD51B
RAD51D
RAD54B
RAD54L2
RAET1E
RAET1E-AS1
RAF1
RAI1
RAI14
RALGAPA2
RALGDS
RALGPS1
RALGPS2
RANBP6
RAP1A
RAP1GAP
RAP1GAP2
RAP2B
RAPGEF1
RAPGEF3
RAPGEF4
RAPGEF6
RAPH1
RARA
RARB
RARG
RARRES1
RARRES3
RASA3
RASA4
RASA4B
RASAL2
RASD2

RASGEF1A
RASGEF1B
RASGRF1
RASGRP2
RASSF2
RASSF3
RASSF4
RASSF6
RASSF9
RB1CC1
RBBP7
RBBP8NL
RBFOX2
RBFOX3
RBKS
RBM18
RBM20
RBM47
RBM7
RBMS1
RBMS2
RBMS3
RBMS3-AS3
RBPMS
RBPMS2
RC3H1
RCAN1
RCAN3
RCC2
RCL1
RCOR1
RCSD1
RD3
RDH10
RDH5
RDX
REEP3
REEP4
REEP5
RELL1
REP15
REPS1
RERE
REREP3
RERG
REXO1
REXO1L2P
RFC5
RFESD
RFFL
RFK
RFT1
RFTN1
RFTN2
RFX2
RFX3
RFX7
RFX8
RGAG1
RGL1
RGMA
RGS12
RGS17
RGS20
RGS3
RGS9
RHBDF2
RHBDL3
RHCG
RHEB
RHOB
RHOBTB1
RHOBTB2
RHOBTB3
RHOH
RHOQ
RHOT1
RHOV
RHPN2
RIC8B
RIIAD1
RILPL1
RIN1
RIN2
RINL
RIOK1
RIPK1
RIPK2
RLF
RLN3
RMDN2
RMDN2-AS1
RMI2
RMND5A
RMST
RNA5S1
RNASE11
RNASE4
RNASET2
RND1
RND3
RNF103-CHMP3
RNF115
RNF121
RNF126
RNF130
RNF141
RNF144A
RNF144B
RNF145
RNF152
RNF157
RNF169
RNF185
RNF19A
RNF212
RNF217
RNF219
RNF220
RNF38
RNF43
RNFT2
RNLS
RNPEPL1
RNU105C
RNU5A-1
RNU5F-1
RNU6-15P
RNU6-2
RNU6-46P
RNU6-67P
RNU6-75P
RNU6ATAC
ROBO1
ROCK1
ROCK2
ROPN1L
ROR1
ROR2
RORA
ROS1
RP2
RPA3-AS1
RPAP1
RPGR
RPGRIP1
RPH3A
RPH3AL
RPL13P5
RPL22
RPL23AP7
RPL3
RPL32P3
RPL36A-HNRNPH2
RPL38
RPRD1B
RPS16
RPS24
RPS3
RPS6KA2
RPS6KB1
RPS6KC1
RPSAP52
RPTN
RPTOR
RRAD
RRAS
RRAS2
RRBP1
RREB1
RRN3P2
RRP1
RRP15
RRP7B
RSF1
RSL1D1
RSPH10B
RSPH10B2
RSPO2
RSPO3
RSPRY1
RTKN2
RTN1
RTN4IP1
RTN4RL1
RTN4RL2
RTTN
RUNX1
RUNX2
RUNX3
RUSC2
RUVBL1
RUVBL2
RWDD3
RXRA
RYBP
RYK
RYR2
S100A10
S100A16
S100A2
S100P
S100Z
S1PR3
S56528
SAA1
SACS
SACS-AS1
SAE1
SAG
SAMD12
SAMD4A
SAMD8
SAP30BP
SAPCD2
SASH1
SAT1
SAV1
SBF2
SBK3
SBNO2
SCAF11
SCAMP2
SCAPER
SCARA5
SCARB1
SCEL
SCFD1
SCG2
SCG5
SCGN
SCHIP1
SCHLAP1
SCIMP
SCMH1
SCN10A
SCN4A
SCNN1A
SCNN1B
SCNN1G
SCOC
SCPEP1
SCRIB
SCRT2
SCTR
SDAD1
SDC1
SDC4
SDC4P
SDCCAG3
SDHA
SDHAP1
SDHAP2
SDK1
SDK2
SDPR
SDR9C7
SDSL
SEC14L1
SEC14L2
SEC16B
SEC22C
SEC24C
SEL1L
SEL1L3
SELENBP1
SELO
SELPLG
SEMA3A
SEMA3C
SEMA3F
SEMA4B
SEMA4D
SEMA5B
SEMA6B
SEPT4
SEPT7P9
SEPT9
SERHL2
SERINC5
SERP2
SERPINA2
SERPINA3
SERPINA4
SERPINA6
SERPINB11
SERPINB12
SERPINB3
SERPINB5
SERPINB6
SERPINE1
SERPINE2
SERPINF1
SERPINF2
SERPINI1
SERTAD3
SESN2
SESTD1
SETBP1
SETD5
SETD7
SETD8
SF3B6
SFI1
SFMBT1
SFMBT2
SFR1
SFTA3
SFTPB
SFXN5
SGCD
SGIP1
SGK1
SGK2
SGK223
SGMS1
SGMS2
SGPP2
SGSM1
SGSM2
SH2B2
SH2D1B
SH2D3C
SH2D4A
SH2D4B
SH2D6
SH2D7
SH3BGRL2
SH3BP4
SH3BP5
SH3D19
SH3GL3
SH3GLB1
SH3KBP1
SH3PXD2A
SH3PXD2B
SH3RF1
SH3RF2
SH3RF3
SH3TC1
SH3TC2
SHANK2
SHB
SHC1
SHC3
SHE
SHH
SHMT1
SHOX
SHQ1
SHROOM3
SIAH1
SIDT1
SIK1
SIK3
SIL1
SIM1
SIMC1
SIN3B
SIPA1L1
SIPA1L2
SIPA1L3
SIRPA
SIRT3
SIX3
SIX3-AS1
SKA3
SKAP1
SKINTL
SLA
SLA2
SLAIN2
SLC10A7
SLC12A3
SLC12A8
SLC13A2
SLC15A1
SLC15A5
SLC16A10
SLC16A12
SLC16A3
SLC16A5
SLC16A6
SLC17A1
SLC17A5
SLC17A9
SLC19A2
SLC19A3
SLC1A2
SLC1A3
SLC1A5
SLC1A7
SLC20A1
SLC20A2
SLC22A1
SLC22A18
SLC22A18AS
SLC22A2
SLC22A23
SLC22A3
SLC22A5
SLC23A1
SLC23A2
SLC23A3
SLC24A3
SLC24A4
SLC25A13
SLC25A18
SLC25A21
SLC25A24
SLC25A30
SLC25A33
SLC25A37
SLC25A45
SLC25A5
SLC26A1
SLC26A11
SLC26A3
SLC26A8
SLC26A9
SLC27A5
SLC28A1
SLC28A3
SLC29A2
SLC29A3
SLC2A1
SLC2A1-AS1
SLC2A14
SLC2A2
SLC2A3
SLC2A5
SLC2A7
SLC2A9
SLC30A1
SLC35D2
SLC35E2
SLC35E2B
SLC35F2
SLC35F3
SLC35F6
SLC35G1
SLC36A4
SLC38A11
SLC38A4
SLC38A6
SLC38A9
SLC39A11
SLC39A14
SLC3A1
SLC3A2
SLC40A1
SLC41A2
SLC41A3
SLC43A1
SLC43A2
SLC44A1
SLC44A2
SLC44A3
SLC44A4
SLC45A1
SLC45A4
SLC47A1
SLC47A2
SLC4A4
SLC4A7
SLC4A8
SLC50A1
SLC51B
SLC5A1
SLC5A10
SLC5A11
SLC5A3
SLC5A4
SLC6A12
SLC6A15
SLC6A17
SLC6A19
SLC6A3
SLC6A6
SLC7A11
SLC7A14
SLC7A2
SLC7A5
SLC7A5P1
SLC7A7
SLC7A8
SLC8A3
SLC9A1
SLC9A2
SLC9A3R2
SLC9A6
SLC9A7
SLC9A7P1
SLC9A8
SLC9B1
SLC9C2
SLCO1B3
SLCO1B7
SLCO1C1
SLCO2A1
SLCO2B1
SLCO4A1
SLCO5A1
SLFN11

SLFN5
SLIT3
SLITRK6
SLMO2
SLMO2-ATP5E
SLN
SLX4IP
SMA4
SMA5
SMAD2
SMAD3
SMAD6
SMAD7
SMAP1
SMARCA1
SMARCD2
SMCO4
SMCR9
SMG1P2
SMG1P5
SMG9
SMIM12
SMIM14
SMIM2
SMIM2-AS1
SMIM21
SMIM24
SMIM3
SMOC1
SMOC2
SMOX
SMPDL3A
SMPDL3B
SMS
SMUG1
SMURF1
SMURF2
SMYD3
SMYD4
SNAI1
SNAP25
SNAP25-AS1
SNAPC3
SNAPC5
SNAR-A12
SNAR-E
SNAR-I
SNED1
SNHG16
SNIP1
SNN
SNORA62
SNORD114-28
SNORD3B-1
SNORD3C
SNORD3D
SNORD56B
SNORD95
SNRNP48
SNTB2
SNX13
SNX24
SNX29
SNX29P2
SNX3
SNX30
SNX33
SNX8
SNX9
SOAT2
SOCS2-AS1
SOCS3
SOCS5
SOCS6
SOGA1
SOHLH1
SON
SORBS1
SORBS2
SORCS2
SORCS3
SOS1
SOS2
SOWAHC
SOX13
SOX17
SOX6
SOX9
SOX9-AS1
SP110
SP3
SP4
SPACA6P
SPAG16
SPAG9
SPARC
SPARCL1
SPATA13
SPATA3
SPATA31D1
SPATA41
SPATA5
SPATC1
SPATS1
SPATS2
SPATS2L
SPC24
SPDEF
SPDYA
SPDYE5
SPDYE7P
SPECC1
SPECC1L
SPECC1L-ADORA2A
SPEG
SPG7
SPIC
SPIDR
SPIN1
SPINK1
SPINK13
SPINK4
SPINK5
SPIRE1
SPNS2
SPOCK1
SPON2
SPPL2B
SPRED2
SPRNP1
SPRY1
SPRY2
SPRY4
SPSB1
SPTB
SPTBN1
SPTBN2
SPTLC2
SQRDL
SRBD1
SRC
SRCAP
SRCIN1
SRCRB4D
SRD5A2
SRD5A3
SREK1
SRGAP1
SRGAP2B
SRGAP2C
SRGAP2D
SRGAP3
SRGN
SRL
SRMS
SRPK2
SRPRB
SRRM1
SRRM2
SRRM3
SRSF12
SRSF3
SRSF5
SRXN1
SSBP2
SSBP3
SSC5D
SSH1
SSH2
SSR2
SSR3
SSR4P1
SSRP1
SSTR4
SSTR5
SSTR5-AS1
SSUH2
ST18
ST20-MTHFS
ST3GAL1
ST3GAL3
ST3GAL4
ST3GAL6
ST5
ST6GAL1
ST6GALNAC1
ST6GALNAC5
ST7
ST7L
ST8SIA3
STAC
STAC2
STAG3L1
STAG3L2
STAM2
STARD13
STARD13-AS
STARD3
STARD4-AS1
STARD8
STAT4
STAT5A
STAT5B
STAU2
STAU2-AS1
STC1
STEAP1
STEAP1B
STEAP2
STEAP3
STIM1
STK11
STK17B
STK24
STK3
STK32B
STK32C
STK35
STK39
STK40
STMN1
STMND1
STOM
STON2
STOX2
STPG1
STPG2-AS1
STRA6
STRBP
STT3B
STX16
STX18-AS1
STX1A
STX8
STXBP1
STXBP5-AS1
STXBP5L
STXBP6
SUCLG2
SUCLG2-AS1
SUCO
SUDS3
SUGCT
SULF2
SULT1B1
SULT1C3
SULT1C4
SULT2B1
SUMF1
SUN2
SUN3
SUSD1
SUSD2
SUSD3
SUSD5
SUV420H1
SUZ12
SUZ12P1
SVIL
SYBU
SYCE1
SYF2
SYK
SYMPK
SYN2
SYN3
SYNE1
SYNE2
SYNGR4
SYNJ2
SYNPO
SYNPO2
SYP
SYP-AS1
SYPL1
SYS1
SYT1
SYT12
SYT13
SYT2
SYT8
SYTL2
SYTL3
SYTL4
SYVN1
SZT2
T
T-Cell Receptor- V-alpha region
T-cell receptor-alpha-chain
TAB2
TAC4
TACC1
TACC2
TACC3
TACR1
TACR2
TAF1B
TAF5L
TAF6
TANC1
TANC2
TANGO6
TANK
TAS1R3
TASP1
TATDN3
TAX1BP1
TBC1D1
TBC1D10A
TBC1D14
TBC1D16
TBC1D2
TBC1D21
TBC1D22A
TBC1D28
TBC1D29
TBC1D2B
TBC1D3
TBC1D31
TBC1D3G
TBC1D3H
TBC1D3P1-DHX40P1
TBC1D3P2
TBC1D4
TBC1D5
TBC1D8
TBC1D9
TBC1D9B
TBCD
TBCE
TBCEL
TBCK
TBL1X
TBX15
TBX19
TBX3
TBX4
TBXA2R
TBXAS1
TCEA1
TCEAL1
TCEB1
TCF12
TCF20
TCF7L2
TCOF1
TCP11L1
TCRA
TCRAV2S1J22
TCRAV9S1
TCRB
TCRBV10S1P
TCRBV19S1P
TCRBV2S1
TCTE1
TDRD9
TEAD1
TEAD2
TEAD4
TECPR1
TECTA
TEF
TEK
TEKT5
TENC1
TENM2
TENM3
TENM4
TERF2
TES
TESC
TESPA1
TEX11
TEX2
TEX26
TEX33
TEX35
TEX41
TF
TFAM
TFAP2A
TFAP2A-AS1
TFAP2C
TFCP2L1
TFDP2
TFEB
TFPI
TG
TGFA
TGFB2
TGFBI
TGFBR1
TGFBR2
TGFBR3
TGIF1
TGM2
TGM3
TGM4
TGS1
THADA
THBS1
THBS2
THEM4
THOC2
THOC3
THRA
THRAP3
THRB
THRB-AS1
THSD4
THSD7A
TIAM1
TIAM2
TICRR
TIGD2
TIGD3
TIMD4
TIMM21
TIMM22
TIMM23
TIMP2
TIMP4
TINAG
TIPARP
TIPARP-AS1
TJP1
TJP3
TK2
TLDC1
TLE1
TLE4
TLK2
TLN1
TLN2
TLR5
TLX1NB
TM2D3
TM4SF1
TM4SF1-AS1
TM4SF18
TM4SF20
TM4SF4
TM4SF5
TMC1
TMC5
TMC7
TMCC1
TMCC1-AS1
TMCO5B
TMED10
TMEM100
TMEM105
TMEM120B
TMEM131
TMEM132B
TMEM143
TMEM151B
TMEM156
TMEM164
TMEM17
TMEM170B
TMEM178B
TMEM183B
TMEM184B
TMEM189
TMEM189-UBE2V1
TMEM191A
TMEM2
TMEM204
TMEM220-AS1
TMEM229B
TMEM230
TMEM236
TMEM243
TMEM245
TMEM254-AS1
TMEM255B
TMEM256
TMEM259
TMEM37
TMEM38A
TMEM41B
TMEM43
TMEM50B
TMEM51
TMEM53
TMEM55A
TMEM56
TMEM56-RWDD3
TMEM60
TMEM61
TMEM63C
TMEM64
TMEM65
TMEM67
TMEM72-AS1
TMEM8C
TMEM91
TMEM99
TMIGD1
TMOD1

TMOD3
TMPPE
TMPRSS6
TMPRSS7
TMSB10
TMSB4X
TMTC2
TMX4
TNFAIP3
TNFAIP8
TNFAIP8L1
TNFAIP8L3
TNFRSF10A
TNFRSF10B
TNFRSF11B
TNFRSF1B
TNFRSF21
TNFRSF4
TNFSF12
TNFSF12-TNFSF13
TNFSF13
TNFSF15
TNFSF18
TNFSF8
TNIK
TNIP1
TNK2
TNKS1BP1
TNNI2
TNNT2
TNPO1
TNRC6A
TNRC6B
TNS1
TNS3
TNS4
TOB2
TOLLIP
TOM1
TOM1L2
TOMM40L
TOMM5
TOP1
TOR1AIP2
TOX
TOX2
TP53BP1
TP53BP2
TP53I3
TP63
TP73
TPCN1
TPCN2
TPD52
TPD52L1
TPGS1
TPM1
TPM2
TPM4
TPO
TPPP
TPRA1
TPRG1
TPTE
TPTE2
TRA
TRABD2B
TRAC
TRAF1
TRAF2
TRAF3
TRAF3IP2
TRAF3IP2-AS1
TRAF4
TRAK1
TRAM1
TRAM2
TRAP1
TRAPPC10
TRAPPC12
TRAPPC3
TRAPPC6A
TRAPPC9
TRERF1
TRERNA1
TRHR
TRIB1
TRIB3
TRIM16
TRIM16L
TRIM2
TRIM26
TRIM27
TRIM29
TRIM35
TRIM47
TRIM48
TRIM49B
TRIM51
TRIM52-AS1
TRIM55
TRIM59
TRIM63
TRIM67
TRIM73
TRIM8
TRIM9
TRIML1
TRIO
TRIOBP
TRIP10
TRIP11
TRIP12
TRIP13
TRIP4
TRMT11
TRMT6
TRNP1
TRPA1
TRPM1
TRPM2
TRPM3
TRPM6
TRPM8
TRPV2
TRPV4
TSC2
TSC22D2
TSC22D3
TSEN15
TSG1
TSHZ2
TSKU
TSN
TSNARE1
TSNAX
TSPAN14
TSPAN15
TSPAN16
TSPAN19
TSPAN2
TSPAN3
TSPAN4
TSPAN5
TSPAN9
TSPEAR
TSPYL2
TSR2
TSSC1
TSTD1
TTC12
TTC16
TTC17
TTC21A
TTC23
TTC23L
TTC26
TTC28
TTC39A
TTC39B
TTC39C
TTC40
TTC5
TTC7A
TTC7B
TTI1
TTI2
TTLL6
TUBA1C
TUBA3C
TUBA3E
TUBB3
TUBD1
TUBGCP3
TUBGCP5
TUBGCP6
TUFT1
TULP4
TUSC5
TUT1
TVP23B
TXN2
TXNDC11
TXNDC16
TXNDC5
TXNIP
TXNL1
TXNL4A
TXNRD1
TXNRD2
TYRO3
TYW1
TYW1B
U2AF2
U2SURP
U4atac
UACA
UBAC1
UBAC2
UBAC2-AS1
UBAP1
UBAP2
UBASH3A
UBASH3B
UBB
UBBP4
UBC
UBE2E1
UBE2E2
UBE2E3
UBE2G2
UBE2H
UBE2K
UBE2Q2L
UBE2Q2P1
UBE2Q2P2
UBE2Q2P3
UBE2QL1
UBE2U
UBE3C
UBL3
UBOX5
UBOX5-AS1
UBR2
UBR3
UBR4
UBXN10
UBXN2B
UBXN6
UCK2
UCN3
UGCG
UGDH
UGDH-AS1
UGT1A1
UGT1A10
UGT1A3
UGT1A4
UGT1A5
UGT1A6
UGT1A7
UGT1A8
UGT1A9
UGT2A3
UGT2B10
UGT2B15
UGT2B4
UGT2B7
UHRF1
UHRF2
ULK1
ULK2
ULK4
UMPS
UNC119B
UNC13B
UNC13C
UNC45B
UNC5A
UNC5B
UNC5C
UNC5CL
UNC93B1
UNG
UNKL
UOX
UPF2
UPK1B
UPK3BL
UPP2
UQCC1
UQCR10
UQCRC1
UQCRHL
URI1
UROC1
USB1
USP10
USP11
USP18
USP2
USP22
USP3
USP30
USP31
USP32
USP32P1
USP34
USP40
USP41
USP43
USP47
USP48
USP7
USP9X
UST
UTP11L
UTRN
UVRAG
VAC14
VAPA
VARS
VARS2
VASN
VASP
VAV1
VAV2
VAV3
VCAN
VCL
VDAC1
VDAC2
VDR
VEGFA
VEGFC
VEPH1
VEZF1
VGF
VGLL3
VGLL4
VIL1
VIMP
VIPR2
VIT
VKORC1L1
VLDLR
VLDLR-AS1
VMP1
VPS13B
VPS13D
VPS37B
VPS53
VRK2
VSIG1
VSIG8
VTI1A
VTRNA2-1
VWA2
VWA5B1
VWA8
VWA8-AS1
VWF
WAPAL
WASH7P
WBP1L
WBSCR16
WBSCR22
WBSCR28
WDFY1
WDFY2
WDR13
WDR18
WDR20
WDR24
WDR25
WDR27
WDR31
WDR35
WDR38
WDR43
WDR45B
WDR60
WDR64
WDR65
WDR70
WDR72
WDR75
WDR81
WDR86
WDR87
WDR93
WDTC1
WEE1
WHAMMP3
WHSC1L1
WIBG
WIPF1
WISP1
WISP2
WNK1
WNK2
WNT16
WNT3
WNT5B
WNT7B
WNT8B
WNT9A
WT1
WWC1
WWC2
WWC3
WWOX
WWP1
WWP2
WWTR1
WWTR1-AS1
X01410
X01411
X04923
X17676
X58749
X61078
X92025
XBP1
XCL1
XIAP
XIRP1
XIRP2
XKRX
XPC
XPO1
XPO6
XPR1
XRCC2
XRCC6BP1
XXYLT1
YAF2
YAP1
YEATS2
YEATS4
YIPF4
YIPF6
YPEL2
YPEL5
YTHDF1
YTHDF3
YWHAG
YWHAQ
Z26593
Z75946
Z77830
ZAK
ZBED1
ZBED3-AS1
ZBTB1
ZBTB12
ZBTB16
ZBTB20
ZBTB24
ZBTB38
ZBTB40
ZBTB7B
ZBTB7C
ZBTB8A
ZBTB8B
ZC2HC1C
ZC3H12A
ZC3H12C
ZC3H14
ZC3H15
ZC3H3
ZC3H7B
ZCCHC14
ZCCHC2
ZCCHC6
ZDHHC11
ZDHHC14
ZDHHC18
ZDHHC20
ZDHHC7
ZDHHC8
ZDHHC8P1
ZDHHC9
ZEB1
ZEB2
ZFAND2A
ZFAND5
ZFAND6
ZFAT
ZFHX3
ZFP36
ZFP36L1
ZFP36L2
ZFX-AS1
ZFYVE20
ZFYVE27
ZHX2
ZHX3
ZIC3
ZMAT4
ZMIZ1
ZMYM2
ZMYND11
ZMYND8
ZNF112
ZNF143
ZNF146
ZNF189
ZNF215
ZNF217
ZNF229
ZNF254
ZNF280C
ZNF280D
ZNF281
ZNF282
ZNF285
ZNF330
ZNF34
ZNF341
ZNF365
ZNF366
ZNF367
ZNF395

ZNF407	ZNF516	ZNF611	ZNF664	ZNF716	ZNF860	ZSWIM5
ZNF414	ZNF532	ZNF618	ZNF664-FAM101A	ZNF721	ZNHIT6	ZSWIM6
ZNF423	ZNF541	ZNF624	ZNF678	ZNF727P	ZNRF1	ZXDA
ZNF438	ZNF57	ZNF628	ZNF70	ZNF765	ZNRF3	ZXDC
ZNF451	ZNF579	ZNF644	ZNF703	ZNF767P	ZNRF3-AS1	ZZEF
ZNF474	ZNF592	ZNF652	ZNF705E	ZNF774	ZNRF4	
ZNF503-AS1	ZNF608	ZNF653	ZNF706	ZNF782	ZPLD1	
ZNF507	ZNF609	ZNF658	ZNF710	ZNF787	ZSCAN25	

List of genes geneset 3

AARS2	C9orf3	DDOST	GAB2	KLF7	MYOCD	PHYH
ABCA13	CA12	DDR2	GAL	KLF9	MYOM1	PIGU
ABCA4	CADPS	DEFB136	GALK2	KRT8	MYT1L	PIK3IP1
ACCSL	CAMK2D	DHX8	GALNT14	KRT80	N4BP2L1	PIK3R1
ACTBL2	CAMKK2	DISP2	GCH1	KRT85	N4BP2L2	PIK3R5
ACTL7B	CAMKMT	DKK3	GCM1	LARS2	NALCN	PIM1
ACYP2	CAPG	DLG2	GFOD1	LAT	NARG2	PIRT
ADAM7	CAPN5	DLX3	GFRA1	LCA5	NDUFS4	PLAU
ADAMTS12	CARD6	DNAJA3	GFRA2	LCN2	NEFL	PLB1
ADCK4	CARTPT	DNASE1L3	GIPC2	LECT1	NELL1	PLCE1
ADD3	CBLN1	DNASE2B	GJA5	LEMD3	NEURL	PLCL2
ADK	CCDC115	DNM3	GMDS	LGI1	NF2	PLEKHA7
ADPRM	CCDC124	DOK1	GMIP	LGI2	NFASC	PLIN2
ADRA2B	CCDC33	DPP4	GNAQ	LHX4	NFATC2	PLSCR4
AGFG1	CCDC58	DPT	GNG10	LIMA1	NFIX	PNKD
AGFG2	CCDC85A	DPYSL3	GNL3L	LIMK1	NFKBIA	POC1A
AHCYL2	CCL11	DRAM1	GOLGA5	LINGO4	NGFR	PODXL
ALK	CCNT2	DRG1	GPR137C	LITAF	NKAIN4	POLDIP3
AMBRA1	CCRN4L	DUSP1	GPR65	LOC152586	NLN	POLM
ANKIB1	CD2	DUSP5	GRB10	LONRF3	NME7	POLR3E
ANKRD6	CD300E	DUSP6	GRIA2	LRBA	NMNAT2	POR
APLNR	CD47	E2F6	GUCY1A3	LRP1B	NOL10	POU2F1
APOBEC3F	CDADC1	EBF3	HAND1	LRRC27	NPHP1	PPAPDC2
APOOL	CDC25A	EDN2	HAPLN2	LRRC28	NPY4R	PPAPDC3
ARHGAP28	CDH22	EEF2K	HARBI1	LRRC7	NR2F2	PPFIA2
ARHGEF28	CDK18	EEFSEC	HBS1L	LRRFIP2	NRG2	PPFIBP2
ARHGEF37	CDK19	EFEMP1	HCN1	LYPLA1	NRG4	PPIL1
ARMC12	CDK5R2	EIF2AK4	HIPK2	LZTS1	NSG1	PPP1R14C
ARNTL	CDK5RAP2	EIF4G2	HIVEP2	MAFB	NSMCE2	PPP1R15B
ARRDC2	CDKL3	ELMOD1	HMX1	MAP1B	NT5DC2	PPP1R1C
ARVCF	CDKN1A	EPAS1	HOXC4	MAP3K1	NTNG1	PRCC
ASB9	CDON	EPB41L3	HPSE2	MAP7	NTSR1	PRKCDBP
ASIC2	CELF3	EPHA5	HS3ST5	MAPK3	NYX	PRL
ASIP	CENPK	EPHB2	HTR1A	MAPKAPK2	OAS1	PRLHR
ASS1	CHAT	ERN1	HTR1D	MARCH10	ODC1	PRR5
ATG7	CHGA	EXT1	ID1	MBOAT2	OLFM1	PRR7
ATOH8	CHGB	EXTL1	ID3	MCAM	ONECUT1	PTGR2
ATP10A	CHRM2	EZH1	IFI44	MCF2L	OPN4	PTK2B
ATP7A	CHST11	F13A1	IFT88	MCOLN2	OPRM1	PTPRZ1
B3GNT2	CLGN	F3	IL20RA	MDH1	OR10D3	RAB3C
B4GALT5	CLK3	FA2H	IL4R	MED10	OR12D2	RAB8B
BAG3	CLPB	FAM103A1	IL6R	MED15	OR13J1	RABGAP1L
BAK1	CLPX	FAM117B	IMPDH1	MED27	OR2C3	RABGGTA
BCAR3	CLVS1	FAM134B	INPP5F	MFGE8	OTUD3	RACGAP1
BCAS3	COLQ	FAM163A	INSC	MGARP	OXR1	RAP1GAP2
BCL2L1	COMMD10	FAM167A	IPCEF1	MINPP1	P2RX2	RAPGEF2
BCL2L10	COMMD7	FAM168A	IRAK2	MIR101-1	P2RY1	RAPGEF5
BCL9L	COQ10B	FAM220A	ISOC1	MIR296	P2RY12	RASGRF1
BCMO1	COTL1	FAM35A	ITGA1	MIR548F5	P4HA2	RCN1
BEGAIN	COX6A2	FAM43A	ITPK1	MKNK2	PAK2	REEP3
BEST3	CREB3L2	FANCC	ITPR1	MMD	PAPOLG	RERE
BLK	CRELD2	FARS2	ITPR2	MNAT1	PARD6B	RGL1
BLOC1S6	CRY2	FARSB	JAKMIP2	MOB2	PARG	RGS5
BNC2	CTDP1	FBLN1	JAM3	MRAP	PARVA	RHOB
BRCA1	CTNNA1	FBXO16	KATNA1	MRPL46	PATZ1	RIIAD1
BRI3BP	CUEDC1	FBXO21	KBTBD11	MRPL48	PC	RIMBP2
BTRC	CX3CR1	FERMT2	KCNA3	MRPS18A	PCBP3	RNF125
BUB1	CXCR2	FGF14	KCNAB1	MRPS27	PCDH12	RNF217
C10orf11	CYB561	FGF2	KCNC1	MRPS30	PDE3A	RNLS
C11orf49	CYB5B	FGFR4	KCNH1	MST4	PDGFC	RPGR
C11orf91	CYP51A1	FILIP1	KCNH2	MT1X	PDHA2	RPH3AL
C1GALT1	CYTH3	FMOD	KCNH6	MTUS1	PDHX	RPP14
C1QTNF1	DAPK2	FNDC7	KCNK9	MUC13	PDK2	RWDD3
C1orf110	DCAF5	FOXK2	KCNQ2	MUM1L1	PDLIM3	S100A6
C1orf21	DCBLD2	FOXO3	KCNQ3	MUSK	PDLIM5	SBF2
C21orf59	DCDC1	FOXP1	KCTD16	MYCBP2	PER1	SBK1
C21orf91	DCPS	FRA10AC1	KIF26B	MYH11	PEX5L	SCD
C5orf15	DDC	FRMD8	KLF13	MYO16	PFDN1	SCG2
C5orf22	DDIT4	FSTL1	KLF14	MYO5B	PHF15	SCN3B

SCN7A
SCUBE1
SDC4
SDK2
SELRC1
SEMA5A
SERP2
SFRP2
SFTPB
SGCD
SGMS1
SH3BP5
SH3GL3
SH3GLB1
SHANK2
SHC4
SIK1
SIL1
SIPA1L1
SKP2
SLAMF9
SLC13A2
SLC16A10
SLC18A1
SLC19A1
SLC1A4
SLC23A1
SLC25A52
SLC29A3
SLC2A9
SLC30A4
SLC36A1
SLC38A8
SLC4A8
SLC6A2
SLCO2B1
SLIT3
SLITRK1
SMAD6
SMARCAL1
SMARCD1
SMOX
SMPD3
SMS
SMYD2
SNAP25
SNAPC1
SND1
SNX7
SORBS3
SORCS1
SORL1
SOST
SPAM1
SPATA9
SPATC1
SPDL1
SPECC1
SPECC1L
SPINK8
SPIRE1
SPP2
SPTB
SPTBN1
SPTLC3
SRP9
SRPK2
SSBP3
SSPN
SSRP1
ST6GALNAC
3
ST7
STAG1
STAT5B
SV2B
SWAP70
SYT1
SYT13
SYT17
TACC2
TAF1B
TBCC
TEAD1
TH
THADA
THRSP
THY1
TLR7
TM4SF20
TMEM2
TMEM252
TMEM27
TMPRSS15
TMTC1
TMX4
TNNC1
TNRC6A
TOM1
TOP1
TPBG
TPGS2
TRAPPC9
TRHDE
TRIM72
TRMT44
TRPM2
TSC22D3
TSPAN1
TSPAN18
TSPYL1
TTC17
TTC8
TUBA3C
TUBB
TUSC5
TVP23C
TWIST1
UACA
UBE2G2
UNC5B
UNC5C
UQCRFS1
URB1
UTP15
VAPB
VAV3
VMP1
VPS13B
WASH2P
WDFY2
WDR92
WRNIP1
WT1
XIRP2
XKR6
XPR1
YPEL5
YWHAH
ZBTB16
ZBTB20
ZC3H12C
ZDHHC13
ZEB1
ZFAT
ZFP36
ZFP90
ZMYND8
ZNF143
ZNF202
ZNF45

Supplementary Table 2 List of genes without SNP data in Generation Scotland cohort from each geneset.

In geneset 1

AGTR2	HTR2C	MECP2	SERPINA7
HDAC6	MAOB	PFKFB1	SFTPA2

In geneset 2

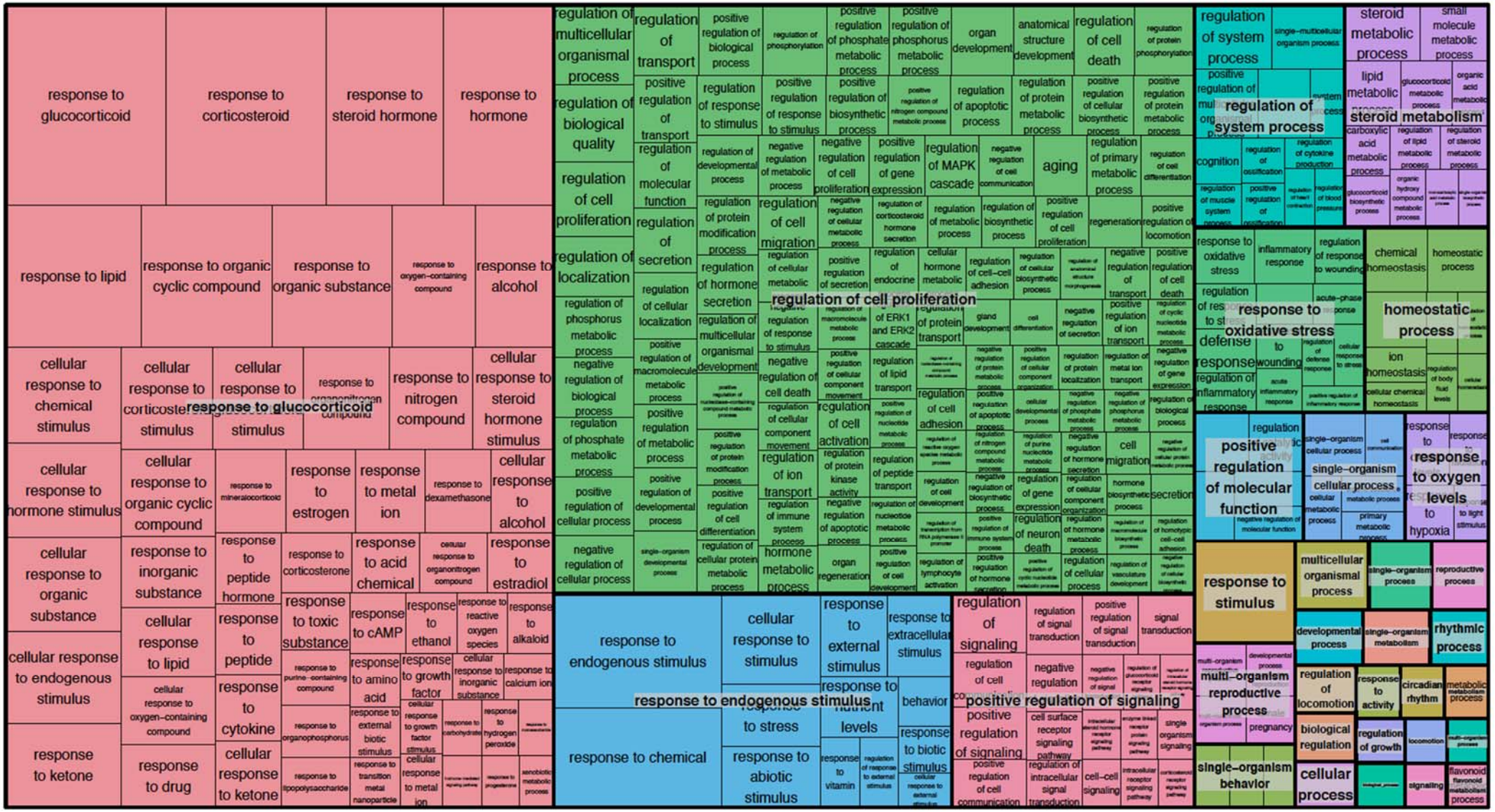
ACRC	DOCK11	HPRT1	NAIP	SDHAP2
ACSL4	DUX4L4	HUWE1	NBEAP1	SEPT7P9
AGAP8	DYNLT3	IGHD	NBPF1	SH3KBP1
AGTR2	E11280	IGSF1	NBPF11	SHOX
AMMECR1	EGFL6	IKBKG	NBPF14	SLC25A5
ANKRD20A2	EIF2S3	IL13RA1	NCF1C	SLC9A6
ANKRD20A3	F8	IL3RA	NDUFA1	SLC9A7
ANKRD30BP2	FAM104B	IQSEC2	NHS	SMA4
APOO	FAM120C	JPX	NOTCH2NL	SMA5
AQP7P1	FAM138A	KIAA1210	NOX1	SMARCA1
AQP7P3	FAM156A	KLHL13	NPIPA3	SMS
AR	FAM156B	KLHL4	NPY4R	SNORD3D
ARAF	FAM223A	LINC00152	NR0B1	SRGAP2B
ARHGAP6	FAM230B	LINC00273	NSDHL	STAG3L2
ARL13A	FAM25G	LINC00842	NUP62CL	STARD8
ARL17B	FAM35BP	LINC00869	NYX	SYP
ARMCX3	FAM35DP	LINC00894	OGT	SYP-AS1
ARSD	FAM47B	LINC00999	OPHN1	SYTL4
ARX	FAM58A	LINC01123	OR2A1	TBC1D3
ASB9	FAM86DP	LINC01138	OR4F16	TBL1X
ASMT	FAM9B	LINC01186	P2RY8	TCEAL1
ATP7A	FAM9C	LINC01203	PCYT1B	TEX11
ATRX	FHL1	LOC100093631	PHEX	THOC2
AWAT1	FIGF	LOC100996255	PHKA1	TMEM164
BAGE	FOXD4L5	LOC101926935	PHKA2-AS1	TMSB4X
BAGE2	FRMPD4	LOC440243	PIN4	TPTE
BAGE5	FTX	LOC440434	PIR-FIGF	TSC22D3
BCOR	FUNDC2	LOC642776	PLAC1	TSPYL2
BEND2	G6PD	LOC643486	PLCXD1	TSR2
C1GALT1C1	GAB3	LOC644172	PLXNA3	UBE2Q2P2
CA5B	GABRE	LOC646743	PMS2L14	UBE2Q2P3
CA5BP1	GATSL2	LOC653501	PMS2P5	UPK3BL
CASK	GEMIN8	LOC729737	POLR2J3	USP11
CD99	GGT3P	LRRC37A3	POM121	USP32P1
CD99L2	GOLGA6L6	LRRc37A	POM121L8P	USP9X
CD99P1	GOLGA8CP	MAGEA4	POTEG	VSIG1
CDKL5	GOLGA8S	MAOA	POTEH	WASH7P
CHEK2P2	GOLGA8T	MAP2K4P1	POTEM	WBSCR16
CHST7	GPR173	MAP7D2	PPEF1	WDR13
CLEC18C	GPR64	MECP2	PPIAL4F	WWC3
COL4A6	GPRIN2	MID1	PRKX	XIAP
CRLF2	GRAPL	MIR1587	PRPS1	XKRX
CTGLF12P	GS1-600G8.3	MIR222	RAB33A	YIPF6
CTPS2	GTF2IP1	MIR3978	RBBP7	ZBED1
CXorf27	GUSBP9	MIR450B	REREP3	ZDHHC9
CXorf56	GYG2	MIR6087	RGAG1	ZFX-AS1
DCAF12L1	HERC2P2	MIR6089-1	RNU6-46P	ZIC3
DDX11L9	HERC2P3	MIR6859-1	RP2	ZNF280C
DDX3X	HERC2P7	MORC4	RPGR	ZNF658
DHRSX	HERC2P9	MORF4L2	RPL36A-HNRNPH2	ZXDA
DIAPH2	HMGB3	MSL3	SAT1	
DLG3	HNRNPH2	MSN		

In geneset 3

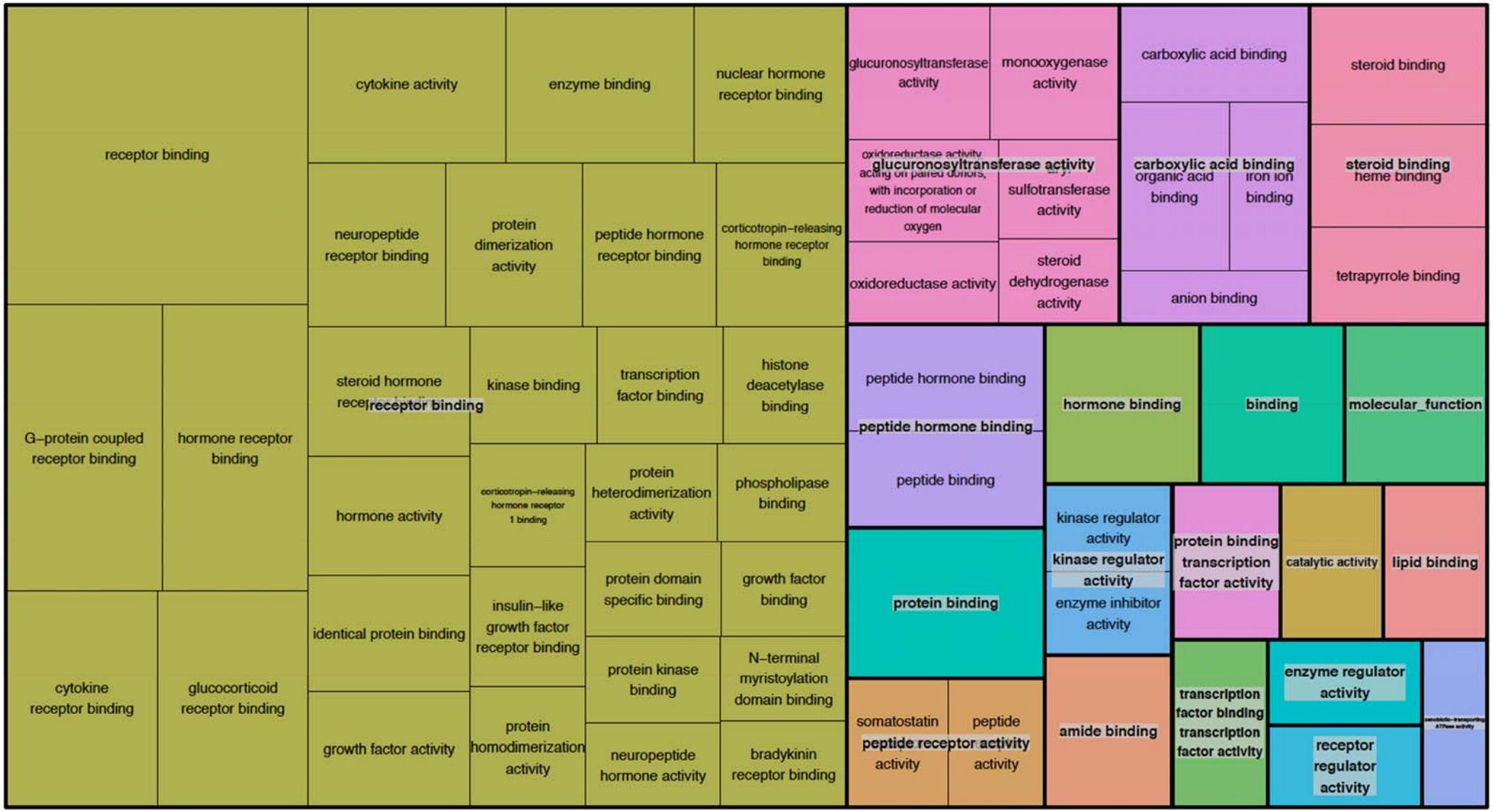
APOOL	GNL3L	MST4	NYX	TLR7
ASB9	HMX1	MUM1L1	RPGR	TMEM27
ATP7A	LONRF3	NPY4R	SMS	TSC22D3

B.2 Supplementary Figures

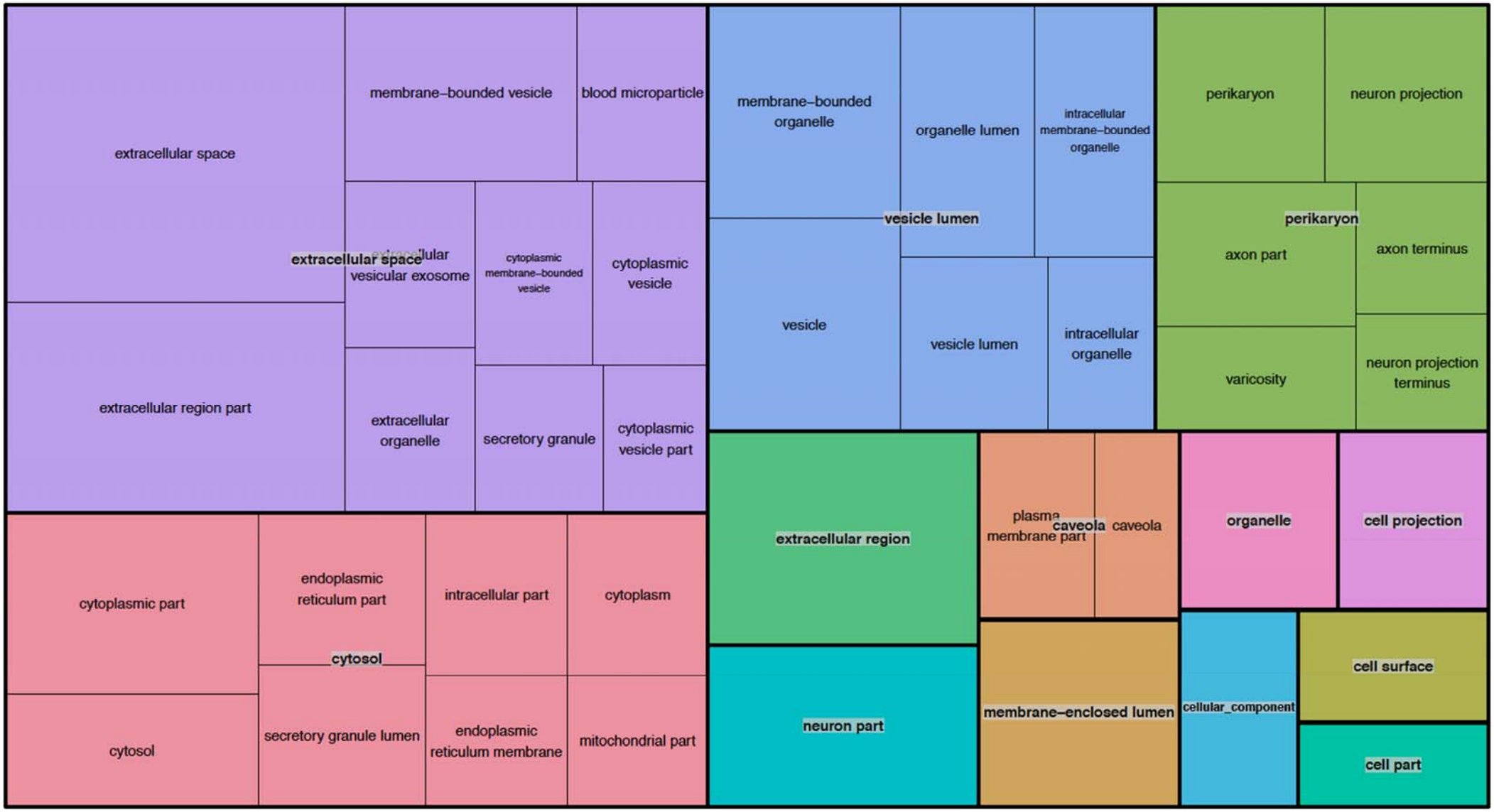
Supplementary Figure 1 Gene Ontology enrichment treemap from Geneset 1: biological processes.



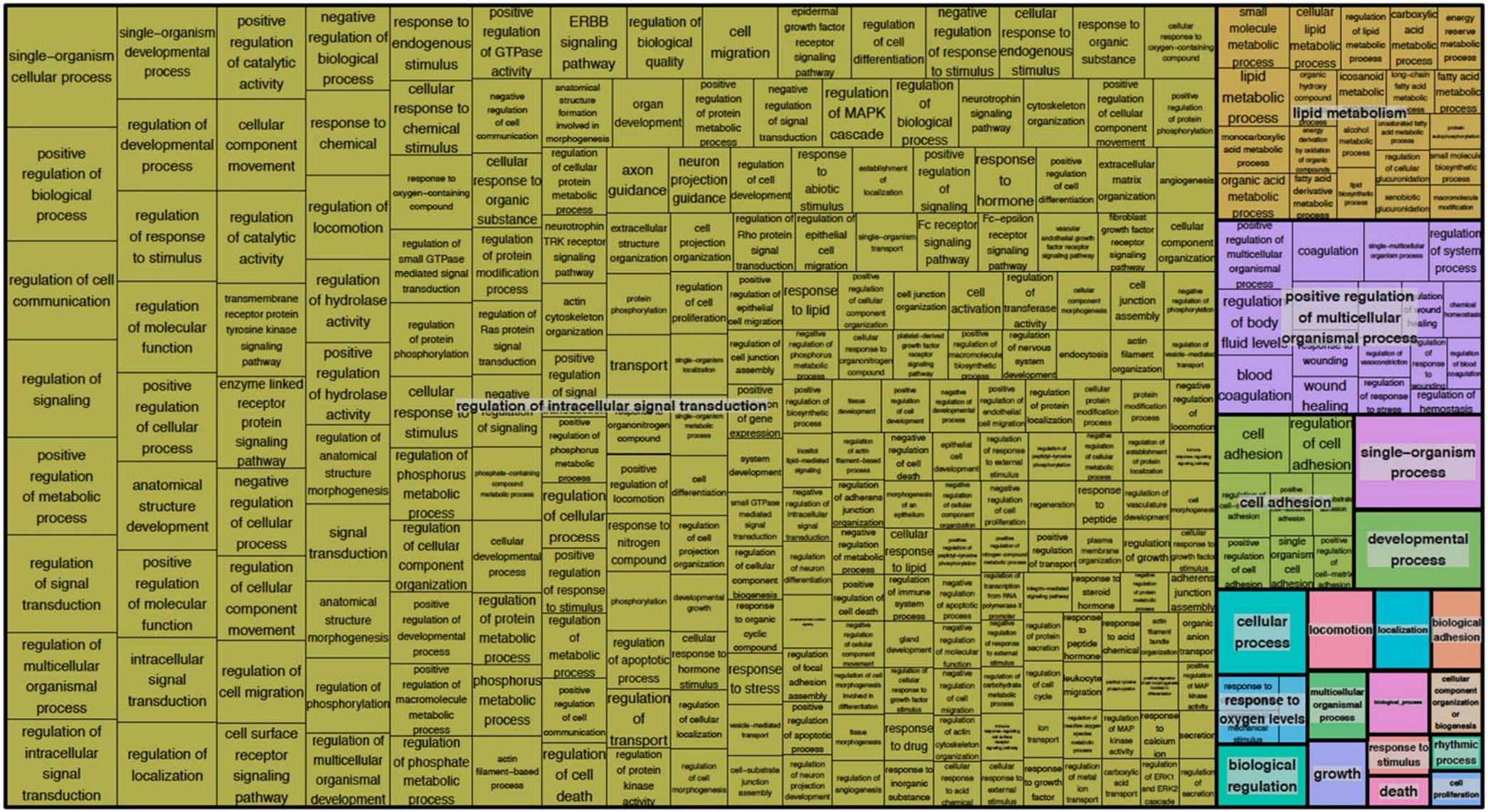
Supplementary Figure 2 Gene Ontology enrichment treemap from Geneset 1: molecular functions.



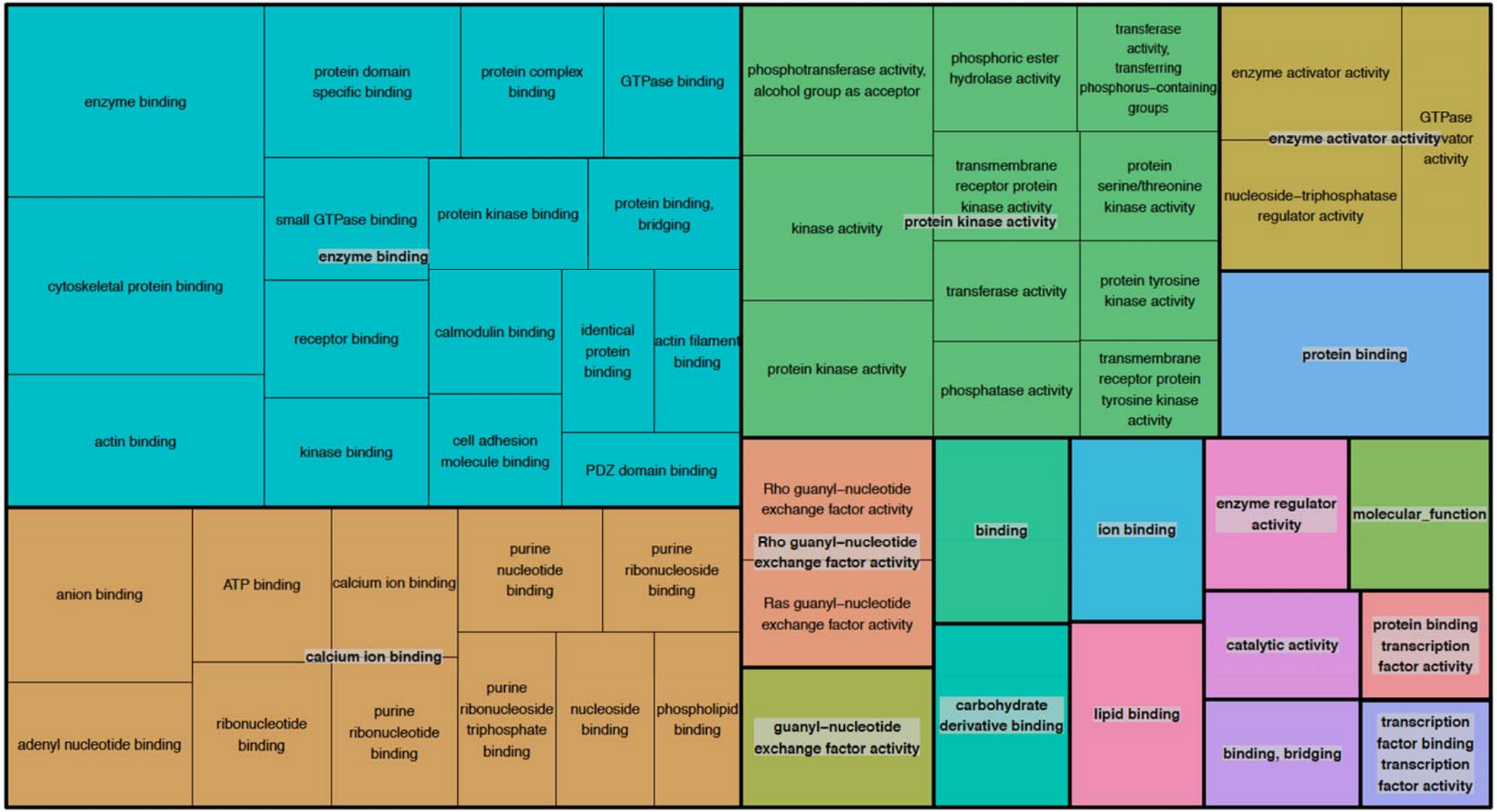
Supplementary Figure 3 Gene Ontology enrichment treemap from Geneset 1: cellular components.



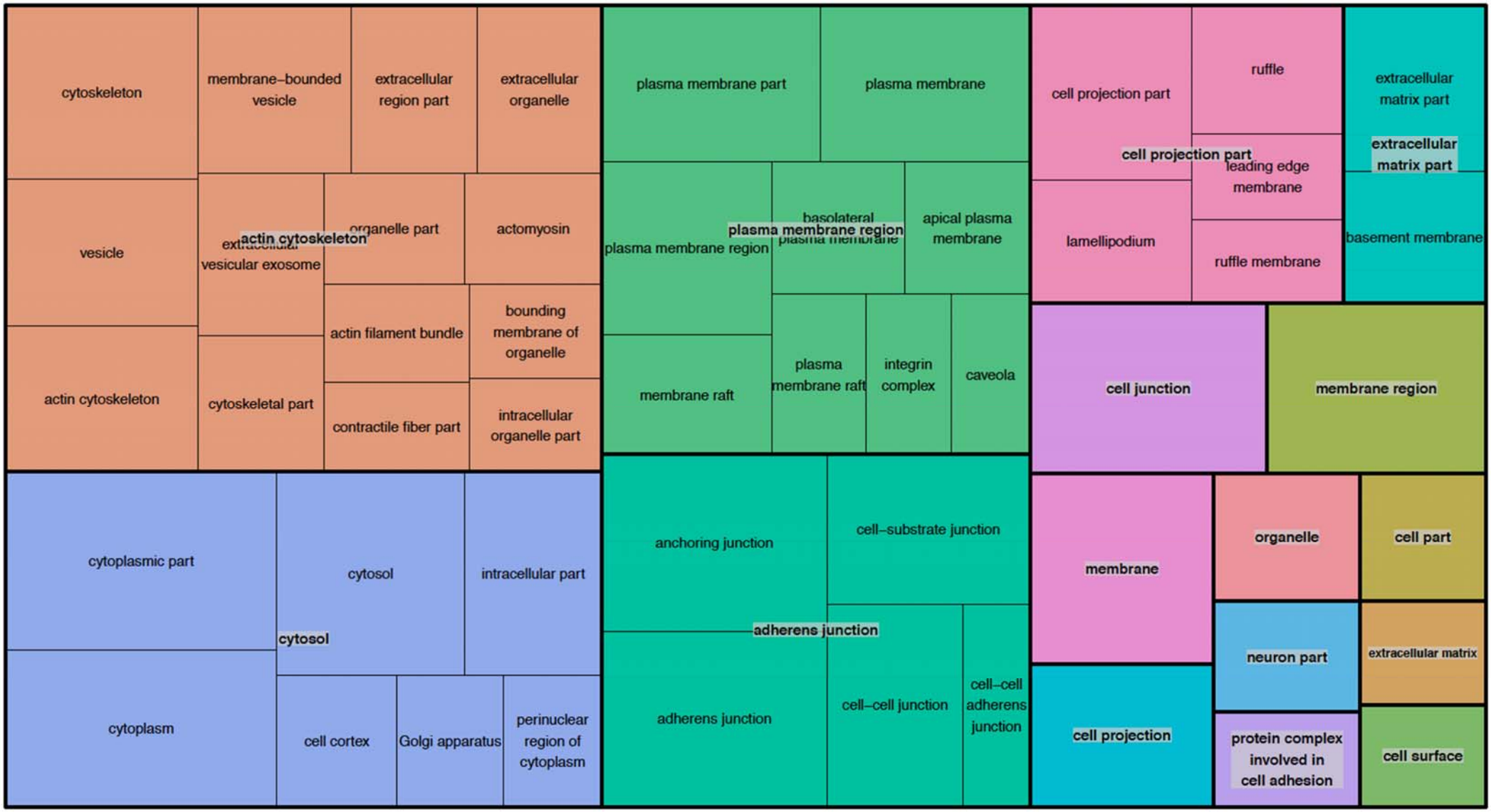
Supplementary Figure 4 Gene Ontology enrichment treemap from Geneset 2: biological processes.



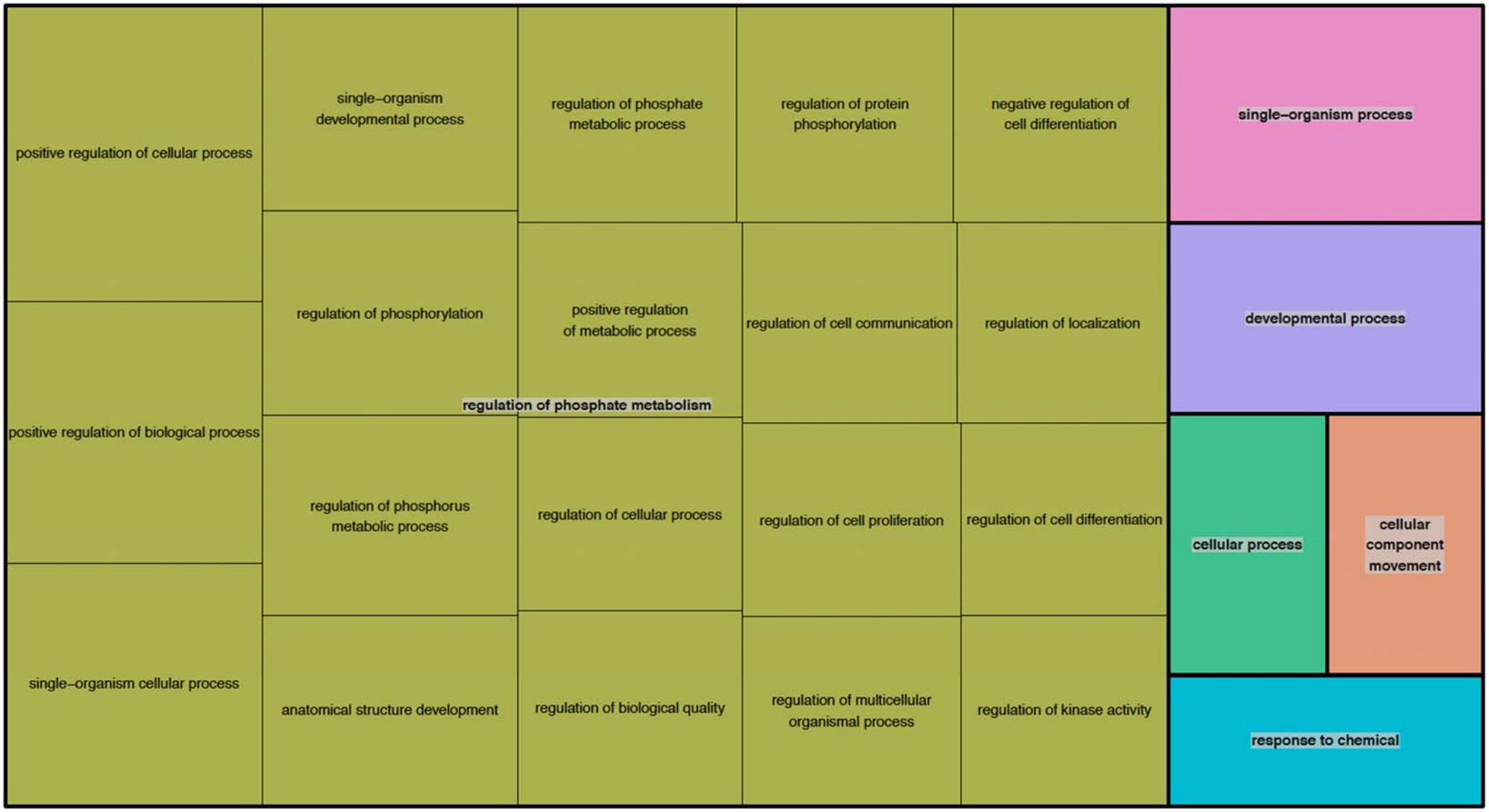
Supplementary Figure 5 Gene Ontology enrichment treemap from Geneset 2: molecular functions.



Supplementary Figure 6 Gene Ontology enrichment treemap from Geneset 2: cellular components.



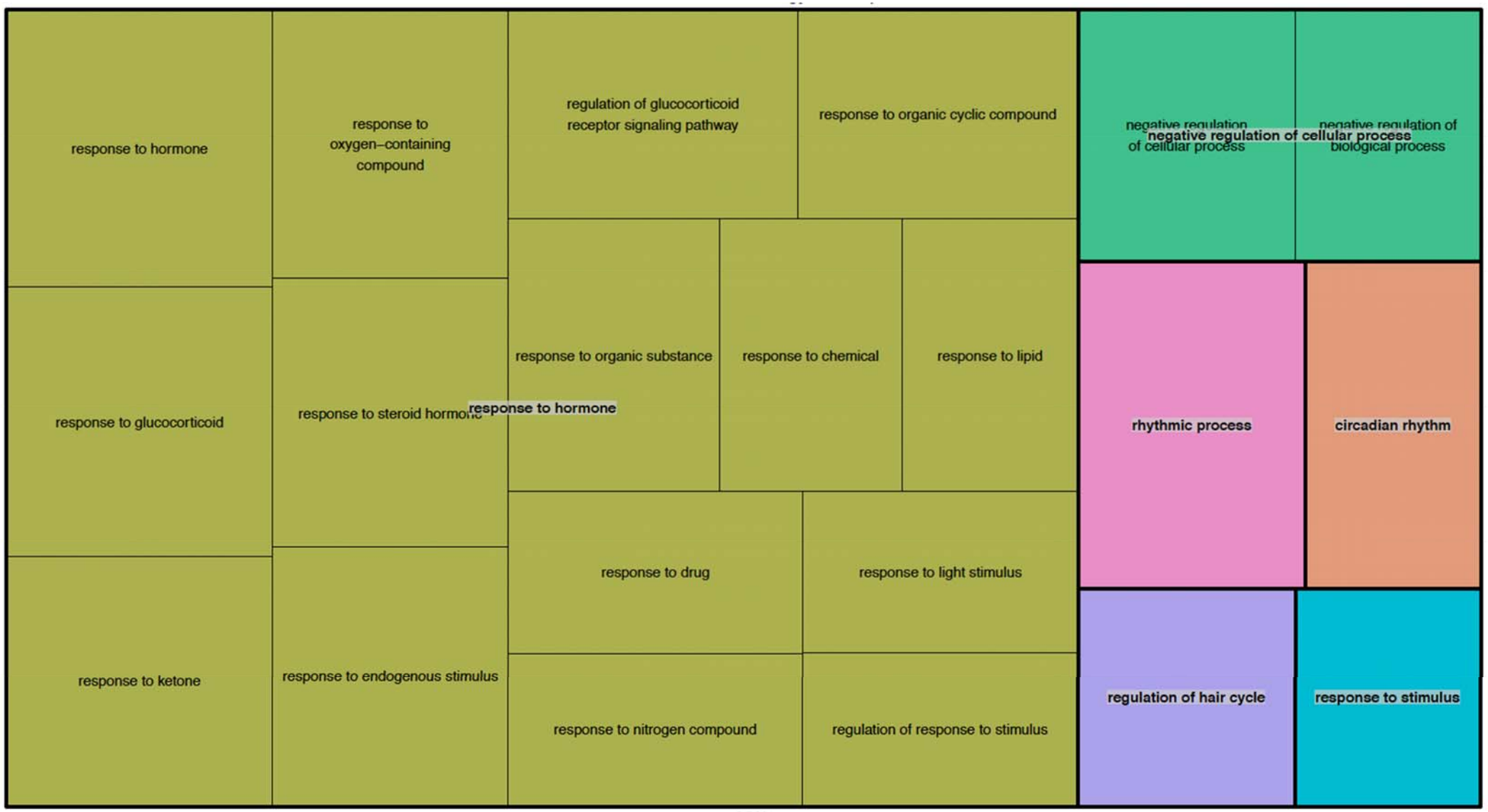
Supplementary Figure 7 Gene Ontology enrichment treemap from Geneset 3: biological processes.



Supplementary Figure 8 Gene Ontology enrichment treemap from Geneset 3: molecular functions.

Supplementary Figure 9 Gene Ontology enrichment treemap from Geneset 3: cellular components.

Supplementary Figure 10 Gene Ontology enrichment treemap for biological processes from genes overlapping across all three glucocorticoid-related genesets.



Appendix C

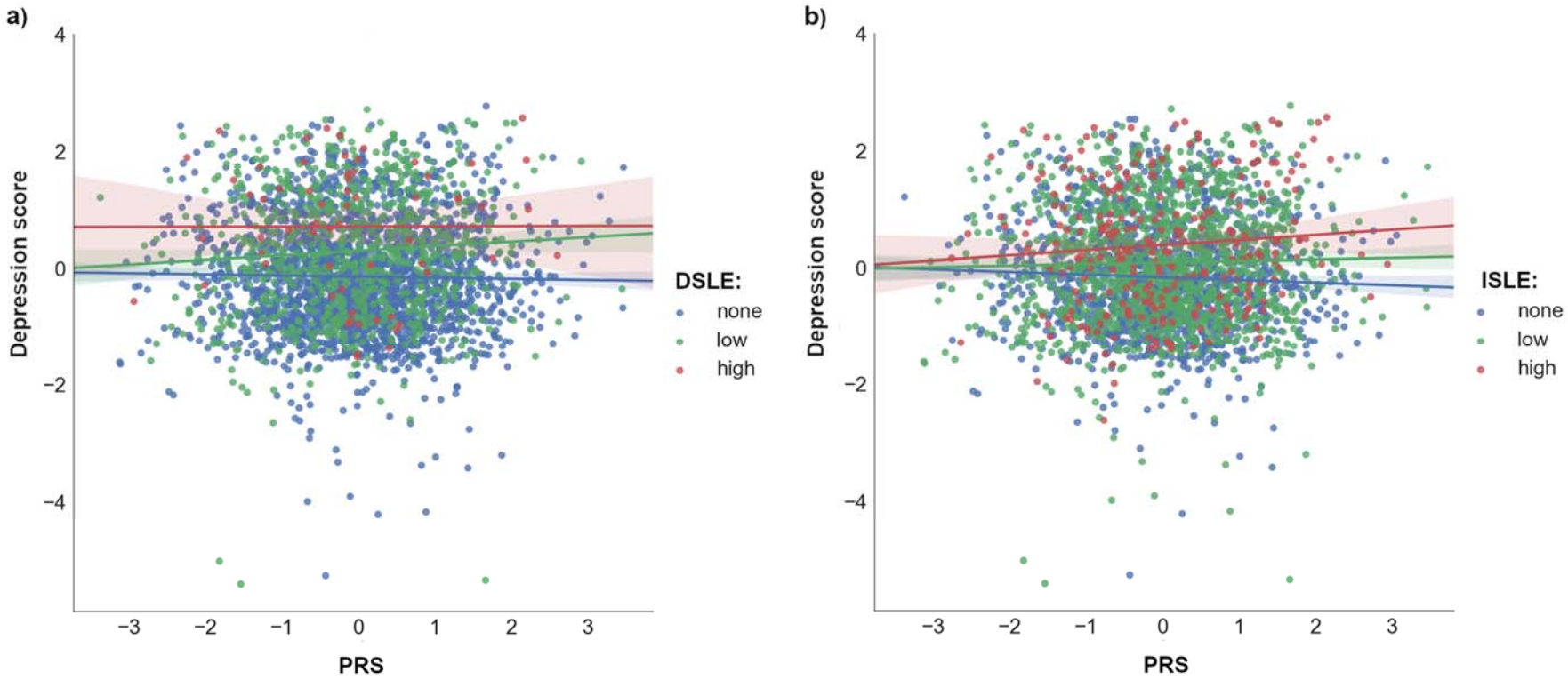
Appendix C contains a supplementary table and supplementary figures for **chapter 4**: A validation of the *diathesis-stress* model for depression in Generation Scotland. The published article in *Translational Psychiatry* is also included.

C.1 Supplementary Table

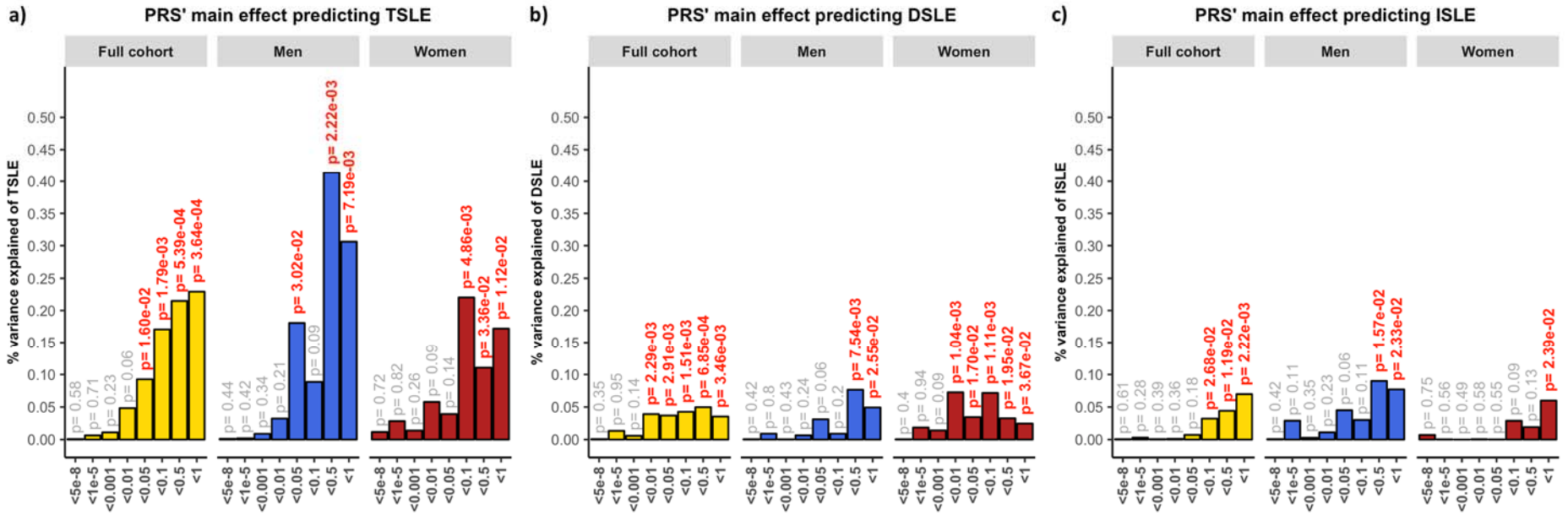
Supplementary Table 1 Summary of SNPs used in PRS profiling. Number of SNPs before clumping and aggregated in PRS after clumping (r^2 = 0.1, within a 10Mb window).

N:	560 351 SNPs	9 411 304 SNPs (imputed)
***p*-value thresholds**	**PRS in Arnau-Soler *et al.***	**PRS in Colodro-Conde *et al.***
$p < 5 \times 10^{-8}$	2	5
$p < 1 \times 10^{-5}$	34	77
$p < 1 \times 10^{-3}$	754	2,044
$p < 0.01$	4,080	11,992
$p < 0.05$	13,749	40,916
$p < 0.1$	22,854	68,444
$p < 0.5$	67,513	204,038
$p < 1$	94,972	280,416

C.2 Supplementary Figures



Supplementary Figure 1 Interaction between PRS and SLE in women. Scatterplot representations of significant *diathesis-stress* interactions on the risk of depressive symptoms in women. X-axis represents the direct effect of PRS (standard deviations from the mean) based on *p*-threshold = $1x10_{-5}$ using **a)** "dependent" SLE (DSLE) or **b)** "independent" SLE (ISLE). Levels of SLE reported by each participant (dot) are categorized in three groups. Blue: 0 SLE, "none"; green: 1 or 2 SLE, "low"; red: 3 or more SLE, "high"; **a)** "none" n = 3 574,"low" n = 1 210,"high" n = 135, **b)** "none" n = 2 369,"low" n = 2 150,"high" n = 394. Y-axis reflects depression score standardized to mean of 0 and standard deviation of 1. Lines represent the increment of risk of depression under a certain category of "stress" dependent on genetic predisposition (= *diathesis*).



Supplementary Figure 2 PRS profile predicting SLE scores. Association between polygenic risk scores (PRS) and **a)** TSLE, **b)** DSLE and **c)** ISLE in STRADL participants. PRS were weighted by summary statistics from the Psychiatric Genetic Consortium MDD GWAS (July 2016), with the exclusion of Generation Scotland participants, constructed at 8 different *p*-thresholds. Full cohort (yellow) was split into men (blue) and women (red). The y-axis represents the proportion of variance in stressful life events (SLE) scores explained by PRS main effects. TSLE: total number of SLE reported based on the List of Threatening Experiences. DSLE: number of SLE reported that are potentially "dependent" on an individual's own behaviour. ISLE: number of potentially "independent" SLE.

C.3 Arnau-Soler *et al.*, 2019, Translational Psychiatry

Arnau-Soler et al. *Translational Psychiatry* (2019)9:25
https://doi.org/10.1038/s41398-018-0356-7

Translational Psychiatry

ARTICLE Open Access

A validation of the diathesis-stress model for depression in Generation Scotland

Aleix Arnau-Soler[1], Mark J. Adams[2], Toni-Kim Clarke[2], Donald J. MacIntyre[2], Keith Milburn[3], Lauren Navrady[2], Generation Scotland, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Caroline Hayward[4], Andrew McIntosh[2,5] and Pippa A. Thomson[1,5]

Abstract

Depression has well-established influences from genetic and environmental risk factors. This has led to the *diathesis-stress* theory, which assumes a multiplicative gene-by-environment interaction (GxE) effect on risk. Recently, *Colodro-Conde et al*. empirically tested this theory, using the polygenic risk score for major depressive disorder (PRS, genes) and stressful life events (SLE, environment) effects on depressive symptoms, identifying significant GxE effects with an additive contribution to liability. We have tested the *diathesis-stress* theory on an independent sample of 4919 individuals. We identified nominally significant positive GxE effects in the full cohort ($R^2 = 0.08\%$, $p = 0.049$) and in women ($R^2 = 0.19\%$, $p = 0.017$), but not in men ($R^2 = 0.15\%$, $p = 0.07$). GxE effects were nominally significant, but only in women, when SLE were split into those in which the respondent plays an active or passive role ($R^2 = 0.15\%$, $p = 0.038$; $R^2 = 0.16\%$, $p = 0.033$, respectively). High PRS increased the risk of depression in participants reporting high numbers of SLE ($p = 2.86 \times 10^{-4}$). However, in those participants who reported no recent SLE, a higher PRS appeared to increase the risk of depressive symptoms in men ($\beta = 0.082$, $p = 0.016$) but had a protective effect in women ($\beta = -0.061$, $p = 0.037$). This difference was nominally significant ($p = 0.017$). Our study reinforces the evidence of additional risk in the aetiology of depression due to GxE effects. However, larger sample sizes are required to robustly validate these findings.

Introduction

Stressful life events (SLE) have been consistently recognized as a determinant of depressive symptoms, with many studies reporting significant associations between SLE and major depressive disorder (MDD)[1–7]. Some studies suggest that severe adversity is present before the onset of illness in over 50% of individuals with depression[8] and may characterize a subtype of cases[9]. However, some individuals facing severe stress never present symptoms of depression[10]. This has led to a suggestion that the interaction between stress and an individual's vulnerability, or *diathesis*, is a key element in the development of depressive symptoms. Such vulnerability can be conceived as a set of biological factors that predispose to illness. Several *diathesis-stress* models have been successfully applied across many psychopathologies[11–15].

The *diathesis-stress* model proposes that a latent *diathesis* may be activated by stress before psychopathological symptoms manifest. Some levels of *diathesis* to illness are present in everybody, with a threshold over which symptoms will appear. Exceeding such a threshold depends on the interaction between *diathesis* and the degree of adversity faced in SLE, which increases the liability to depression beyond the combined additive effects of the *diathesis* and stress alone[11]. Genetic risk factors can, therefore, be conceived as a genetic *diathesis*. Thus, this genetically driven effect produced by the *diathesis-*

Correspondence: Aleix Arnau-Soler (aleix.arnau.soler@igmm.ed.ac.uk) or Pippa A. Thomson (Pippa.Thomson@ed.ac.uk)
[1]Medical Genetics Section, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[2]Division of Psychiatry, Deanery of Clinical Sciences, Royal Edinburgh Hospital, University of Edinburgh, Morningside Park, Edinburgh EH10 5HF, UK
Full list of author information is available at the end of the article.

stress interaction can be seen as a gene-by-environment interaction (GxE).

MDD is characterized by a highly polygenic architecture, composed of common variants with small effect and/or rare variants[16]. Therefore, interactions in depression are also expected to be highly polygenic. In recent years, with the increasing success of genome-wide association studies, GxE studies in depression have shifted towards hypothesis-free genome-wide and polygenic approaches that capture liability to depression using genetic data[17–23,24]. Recent advances in genomics and the massive effort from national institutions to collect genetic, clinical and environmental data on large population-based samples now provide an opportunity to empirically test the *diathesis-stress* model for depression. The construction of polygenic risk scores (PRS) offers a novel paradigm to quantify genetic *diathesis* into a single genetic measure, allowing us to study GxE effects with more predictive power than any single variant[25–28]. PRS are genetic indicators of the aggregated number of risk alleles carried by an individual weighted by their allelic effect estimated from genome-wide association studies. This polygenic approach to assessing the *diathesis-stress* model for depression has been tested using either childhood trauma[17,19,24] or adult SLE[18,23,24] as measures of environmental adversity.

Recently, Colodro-Conde et al.[23] provided a direct test of the *diathesis-stress* model for recent SLE and depressive symptoms. In this study, Colodro-Conde et al. used PRS weighted by the most recent genome-wide meta-analysis conducted by the Psychiatric Genetics Consortium (PGC; $N = 159{,}601$), and measures of three environmental exposures: lack of social support, "personal" SLE, and "network" SLE. Colodro-Conde et al. reported a significant additive risk on liability to depression due to a GxE effect in individuals who combine a high genetic predisposition to MDD and a high number of reported "personal" SLE, mainly driven by effects in women. A significant effect of interaction was not detected in males. They found no significant interaction between the genetic *diathesis* and "network" SLE or social support. They concluded that the effect of stress on risk of depression was dependent on an individual's *diathesis*, thus supporting the *diathesis-stress* theory. In addition, they suggested possible sex-specific differences in the aetiology of depression. However, Colodro-Conde et al. findings have not, to our knowledge, been independently validated.

In the present study, we aim to test the *diathesis-stress* model in an independent sample of 4919 unrelated white British participants from a further longitudinal follow-up from Generation Scotland, and assess the differences between women and men, using self-reported depressive symptoms and recent SLE.

Materials and methods

Sample description

Generation Scotland is a family-based population cohort recruited throughout Scotland by a cross-disciplinary collaboration of Scottish medical schools and the National Health Service (NHS) between 2006 and 2011[29]. At baseline, blood and salivary DNA samples from Generation Scotland participants were collected, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. Genome-wide genotype data were generated using the Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry[30]. The procedures and further details for DNA extraction and genotyping have been extensively described elsewhere[31,32]. In 2014, 21,525 participants from Generation Scotland eligible for re-contact were sent self-reported questionnaires as part of a further longitudinal assessment funded by a Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL)[33] to collect new and updated mental health questionnaires including psychiatric symptoms and SLE measures. 9618 re-contacted participants from Generation Scotland agreed to provide new measures to the mental health follow-up[33] (44.7% response rate). Duplicate samples, those showing sex discrepancies with phenotypic data, or that had more than 2% missing genotype data, were removed from the sample, as were samples identified as population outliers in principal component analysis (mainly non-Caucasians and Italian ancestry subgroups). In addition, individuals with diagnoses of bipolar disorder, or with missing SLE data, were excluded from the analyses. SNPs with more than 2% of genotypes missing, Hardy-Weinberg Equilibrium test $p < 1 \times 10^{-6}$, or a minor allele frequency lower than 1%, were excluded. Individuals were then filtered by degree of relatedness (pi-hat < 0.05) using PLINK v1.9[34], maximizing retention of those participants reporting higher numbers of SLE (see phenotype assessment below). After quality control, the final dataset comprised 4919 unrelated individuals of European ancestry and 560 351 SNPs (mean age at questionnaire: 57.2, s.d. = 12.2, range 22–95; *women*: $n = 2990$–60.8%, mean age 56.1, s.d. = 12.4; *men*: $n = 1\ 929$–39.2%, mean age 58.7, s.d. = 11.8). Further details on the recruitment procedure and Generation Scotland profile are described in detail elsewhere[29,31,35–37]. All participants provided written consent. All components of Generation Scotland and STRADL obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the National Health Service (reference 05/s1401/89). Generation Scotland data is available to researchers on application to the Generation Scotland Access Committee (access@generationscotland.org).

Phenotype assessment

Participant self-reported current depressive symptoms through the 28-item scaled version of The General Health Questionnaire[38,39]. The General Health Questionnaire is a reliable and validated psychometric screening tool to detect common psychiatric and non-psychotic conditions (General Health Questionnaire Cronbach alpha coefficient: 0.82–0.86)[40]. This consists of 28 items designed to identify whether an individual's current mental state has changed over the last 2 weeks from their typical state. The questionnaire captures core symptoms of depression through subscales for severe depression, emotional (e.g., anxiety and social dysfunction) and somatic symptoms linked to depression. These subscales are highly correlated[41] and suggest an overall general factor of depression[42]. Participants rated the 28 items on a four-point Likert scale from 0 to 3 to assess its degree or severity[40] (e.g., *Have you recently felt that life is entirely hopeless?* "Not at all", "No more than usual", "Rather more than usual", "Much more than usual"), resulting on an 84-point scale depression score. The Likert scale, which provides a wider and smoother distribution[40], may be more sensitive to detect changes in mental status in those participants with chronic conditions or chronic stress who may feel their current symptoms as "usual"[43], and to detect psychopathology changes as response to stress. The final depression score was log transformed to reduce the effect of positive skew and provide a better approximation to a normal distribution. In addition, participants completed the Composite International Diagnostic Interview–Short Form, which diagnoses lifetime history of MDD according to DSM-IV criteria[44]. The depression score predicted lifetime history of MDD (odd ratio = 1.91, 95% confidence intervals 1.80–2.02, $p = 1.55 \times 10^{-102}$, $N = 8994$), with a 3.8-fold increased odds of having a lifetime history of MDD between participants in the top and bottom deciles, thus supporting the usefulness of the depression score in understanding MDD. To improve interpretation, we scaled the depression score to a mean of 0 when required (Fig. 3).

Data from a self-reported questionnaire based on the List of Threating Experiences[45] was used to construct a measure of common SLE over the previous 6 months. The List of Threatening Experiences is a reliable psychometric device to measure psychological "stress"[46,47]. It consists of a 12-item questionnaire to assess SLE with considerable long-term contextual effects (e.g., *Over last 6 months, did you have a serious problem with a close friend, neighbor or relatives?*). A final score reflecting the total number of SLE (TSLE) ranging from 0 to 12 was constructed by summing the "yes" responses. Additionally, TSLE was split into two categories based on those items measuring SLE in which the individual may play and active role exposure to SLE, and therefore in which the SLE is influenced by genetic factors and thus subject to be "dependent" on an individual's own behavior or symptoms (DSLE; 6 items, e.g., *a serious problem with a close friend, neighbor or relatives* may be subject to a respondent's own behavior), or SLE that are not influenced by genetic factors, likely to be "independent" on a participant's own behavior (ISLE; 5 items, e.g., *a serious illness, injury or assault happening to a close relative* is potentially independent of a respondent's own behavior)[45,48]. The item "*Did you/your wife or partner give birth?*" was excluded from this categorization. In addition, SLE reported were categorized to investigate the *diathesis* effect at different levels of exposure, including a group to test the *diathesis* effect when SLE is not reported. Three levels of SLE reported were defined (0 SLE = "none", 1 or 2 SLE = "low", and 3 or more SLE = "high") to retain a large enough sample size for each group to allow meaningful statistical comparison.

Polygenic profiling and statistical analysis

Polygenic risk scores (PRS) were generated by PRSice[49], whose functionality relies mostly on PLINK v1.9[34], and were calculated using the genotype data of Generation Scotland participants (i.e., target sample) and summary statistics for MDD from the PGC-MDD2 GWAS release (July 2016, discovery sample) used by Colodro-Conde et al.[23], with the added contribution from QIMR cohort and the exclusion of Generation Scotland participants, resulting in summary statistics for MDD derived from a sample of 50,455 cases and 105,411 controls.

Briefly, PRSice removed strand-ambiguous SNPs and clump-based pruned ($r^2 = 0.1$, within a 10 Mb window) our target sample to obtain the most significant independent SNPs in approximate linkage equilibrium. Independent risk alleles were then weighted by the allelic effect sizes estimated in the independent discovery sample and aggregated into PRS. PRS were generated for eight p thresholds (p thresholds: $< 5 \times 10^{-8}$, $< 1 \times 10^{-5}$, < 0.001, < 0.01, < 0.05, < 0.1, < 0.5, $<= 1$) determined by the discovery sample and standardized (See Supplementary Table 1 for summary of PRS).

A genetic relationship matrix (GRM) was calculated for each dataset (i.e., *full cohort*, *women*, and *men*) using GCTA 1.26.0[50]. Mixed linear models using the GRM were used to estimate the variance in depression score explained by PRS, SLEs and their interaction; and stratified by sex. Twenty principal components were calculated for the datasets.

The mixed linear model used to assess the effects of PRS is as follows:

$$Depression = \beta_0 + \beta_1 PRS + GRM + Covariates$$

Mixed linear models used to assess the effect of the stressors are as follows:

$$Depression = \beta_0 + \beta_1 TSLE + GRM + Covariates$$
$$Depression = \beta_0 + \beta_1 DSLE + GRM + Covariates$$
$$Depression = \beta_0 + \beta_1 ISLE + GRM + Covariates$$

Following Colodro-Conde et al.[23], covariates (i.e., age, age^2, sex, age-by-sex and age^2-by-sex interactions, and 20 principal components) were regressed from PRS (PRS') and SLE scores (i.e., TSLE', DSLE' and ISLE'; SLEs') before fitting models in GCTA to guard against confounding influences on the PRS-by-SLEs interactions[51]. PRS' and SLEs' were standardized to a mean of 0 and a standard deviation of 1. The Mixed linear models (i.e., the *diathesis-stress* model) used to assess GxE effects are as follows:

$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 TSLE' + \beta_3 PRS' xTSLE' + GRM + Covariates$$
$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 DSLE' + \beta_3 PRS' xDSLE' + GRM + Covariates$$
$$Depression = \beta_0 + \beta_1 PRS' + \beta_2 ISLE' + \beta_3 PRS' xISLE' + GRM + Covariates$$

Covariates fitted in the models above were age, age^2, sex, age-by-sex, age^2-by-sex, and 20 principal components. Sex and its interactions (age-by-sex and age^2-by-sex) were omitted from the covariates when stratifying by sex. All parameters from the models were estimated using GCTA and the significance of the effect (β) from fixed effects assessed using a Wald test. The significance of main effects (PRS and SLEs) allowed for nominally testing the significance of interactions at p-threshold = 0.05. To account for multiple testing correction, a Bonferroni's adjustment correcting for 8 PRS and 3 measures of SLE tested (24 tests) was used to establish a robust threshold for significance at $p = 2.08 \times 10^{-3}$.

The PRS effect on depression score at different levels of exposure was further examined for the detected nominally significant interactions by categorizing participants on three groups based on the number of SLE reported (i.e., "none", "low" or "high"). Using linear regression, we applied a least squares approach to assess PRS' effects on the depression score in each SLE category. Further conservative Bonferroni correction to adjust for the 3 SLE categories tested established a threshold for significance of $p = 6.94 \times 10^{-4}$.

Differences on the estimated size of GxE effect between women and men were assessed by comparing a z-score to the standard normal distribution ($\alpha = 0.05$, one-tailed). Z-scores were derived from GxE estimates (β) and standard errors (SE) detected in women and men as follows:

$$Z - score = \frac{\beta_{women} - \beta_{men}}{\sqrt{SE(\beta_{women})^2 + SE(\beta_{men})^2}}$$

Results

PRS for MDD significantly predicted the depression score across the whole sample ($\beta = 0.080$, s.e. = 0.014, $p = 7.53 \times 10^{-9}$) explaining 0.64% of the variance at its best p-threshold (p-threshold = 0.1; Fig. 1a). Stratifying by sex, PRS significantly predicted the depression score in both sexes, explaining 0.59% in men and 0.67% in women (*men*: p-threshold = 0.1, $\beta = 0.077$, s.e. = 0.022, $p = 2.09 \times 10^{-4}$; *women*: p-threshold = 0.1, $\beta = 0.082$, s.e. = 0.018, $p = 4.93 \times 10^{-6}$; Fig. 1a). Self-reported SLE over the last 6 months (TSLE, mean = 1.3 SLE, s.d. = 1.5) also significantly predicted depression score for the whole sample and stratified by sex (*full cohort*: variance explained = 4.91%, $\beta = 0.222$, s.e. = 0.014, $p = 9.98 \times 10^{-59}$; *men*: 4.19%, $\beta = 0.205$, s.e. = 0.021, $p = 2.23 \times 10^{-22}$; *women*: 5.33%, $\beta = 0.231$, s.e. = 0.018, $p = 7.48 \times 10^{-38}$; Fig. 1b). Overall, significant additive contributions from genetics and SLE to depression score were detected in all participants and across sexes. There was no significant difference in the direct effect of TSLE between women and men ($p = 0.17$). However, the variance in depression score explained by the TSLE appeared to be lower than the variance explained by the measure of personal SLE (PSLE) used in Colodro-Conde et al.[23] (12.9%). This may, in part, be explained by different contributions of dependent and independent SLE items screened in Colodro-Conde et al. compared to our study. Although questions about dependent SLE (DSLE, mean = 0.4 SLE) represented over 28% of the TSLE-items reported in our study, the main effect of DSLE explained approximately 93% of the amount of variance explained by TSLE (*full cohort*: variance explained = 4.56%, $\beta = 0.212$, s.e. = 0.014, $p = 1.73 \times 10^{-54}$; *men*: 3.74%, $\beta = 0.193$, s.e. = 0.021, $p = 9.66 \times 10^{-21}$; *women*: 5.07%, $\beta = 0.225$, s.e. = 0.018, $p = 8.09 \times 10^{-35}$; Fig. 1b). Independent SLE (ISLE, mean = 0.85 SLE), which represented over 69% of TSLE-items, explained approximately 57% of the amount of variance explained by TSLE (*full cohort*: variance explained = 2.80%, $\beta = 0.167$, s.e. = 0.014, $p = 1.32 \times 10^{-33}$; *men*: 2.44%, $\beta = 0.156$, s.e. = 0.022, $p = 2.88 \times 10^{-13}$; *women*: 3.02%, $\beta = 0.174$, s.e. = 0.018, $p = 5.20 \times 10^{-22}$; Fig. 1b). To explore the contribution from each measure, we combined DSLE and ISLE together in a single model. DSLE explained 3.34% of the variance in depression score compared to 1.45% of the variance being

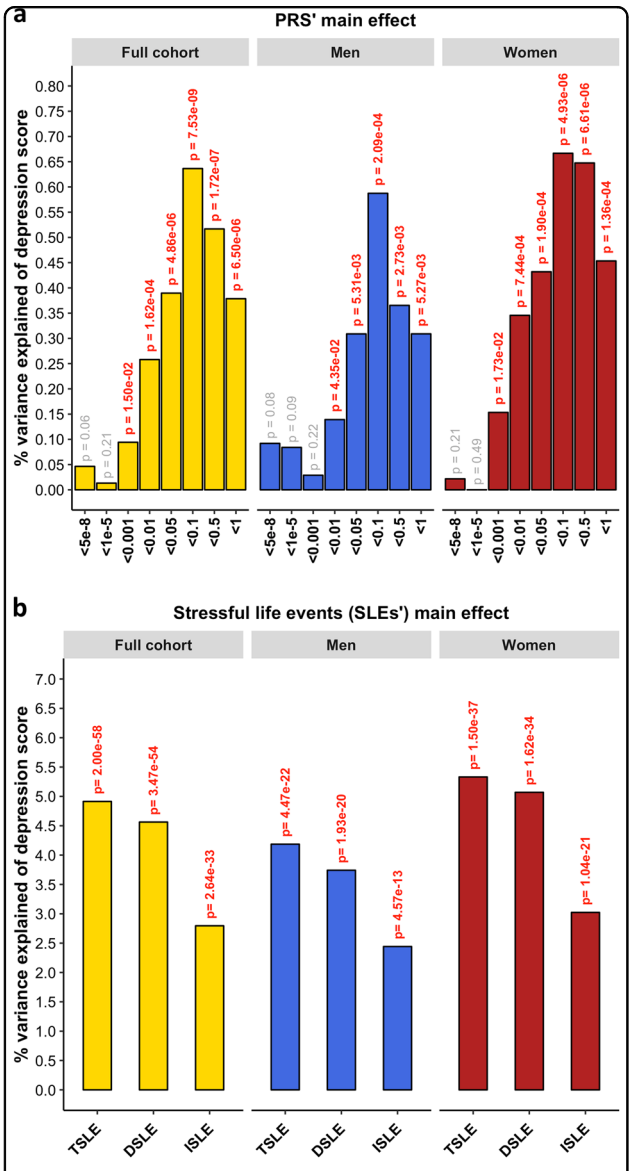


Fig. 1 Prediction of depression symptoms and SLE using the PRS for MDD. **a** Association between polygenic risk scores (PRS) and depression score (main effects, one-sided tests). PRS were generated at 8 *p*-threshold levels using summary statistics from the Psychiatric Genetic Consortium MDD GWAS (released July 2016) with the exclusion of Generation Scotland participants. The depression score was derived from The General Health Questionnaire. The *Y*-axis represents the % of variance of depression score explained by PRS main effects. The full cohort (yellow) was split into men (blue) and women (red). In Colodro-Conde et al. PRS for MDD significantly explained up to 0.46% of depression score in their sample (~0.39% in women and ~0.70% in men). **b** Association between reported number of SLE and depression score (main effect, one-sided tests, results expressed in % of variance in depression score explained). SLE were self-reported through a brief life-events questionnaire based on the List of Threatening Experiences and categorized into: total number of SLE reported (TSLE), "dependent" SLE (DSLE) or "independent" SLE (ISLE). The full cohort (yellow) was split into men (blue) and women (red). In Colodro-Conde et al. "personal" SLE significantly explained up to 12.9% of depression score variance in their sample (~11.5% in women and ~16% in men)[23]

explained by ISLE, suggesting that DSLE have a greater effect on liability to depressive symptoms than ISLE.

A *diathesis-stress* model for depression was tested to assess GxE effects. We detected significant, albeit weak, GxE effects on depression score (Fig. 2). The PRS interaction with TSLE was nominally significant in the full cohort ($\beta = 0.028$, s.e. $= 0.014$, $R^2 = 0.08\%$, $p = 0.049$) and slightly stronger in women ($\beta = 0.044$, s.e. $= 0.018$, $R^2 = 0.19\%$, $p = 0.017$; Fig. 2a), compared to men in which the effect was not significant ($\beta = 0.039$, s.e. $= 0.022$, $R^2 = 0.15\%$, $p = 0.07$). However, these results did not survive correction for multiple testing ($p > 2.08 \times 10^{-3}$).

The best-fit threshold was much lower in women (p-threshold $= 1 \times 10^{-5}$) compared to the full sample (p-threshold $= 0.01$). The size of the GxE effects across sexes at p-threshold $= 1 \times 10^{-5}$ were significantly different (GxE*sex $p = 0.017$), but not at the best p-threshold in the full cohort (p-threshold $= 0.01$, GxE*sex $p = 0.32$; Fig. 2a). In women, GxE effect with DSLE predicted depression score (p-threshold $= 1 \times 10^{-5}$; $\beta = 0.039$, s.e. $= 0.019$, $R^2 = 0.15\%$, $p = 0.038$; Fig. 2b and Supplementary Fig. 2a), as did the GxE effect with ISLE (p-threshold $= 1 \times 10^{-5}$; $\beta = 0.040$, s.e. $= 0.019$, $R^2 = 0.16\%$, $p = 0.033$; Fig. 2c and Supplementary Fig. 2b). No significant interaction was detected in men (best-fit p-threshold $= 0.1$) with either TSLE ($\beta = 0.039$, s.e. $= 0.022$, $R^2 = 0.15\%$, $p = 0.072$; Fig. 2a), DSLE ($\beta = 0.024$, s.e. $= 0.022$, $R^2 = 0.06\%$, $p = 0.28$; Fig. 2b) or ISLE ($\beta = 0.043$, s.e. $= 0.022$, $R^2 = 0.18\%$, $p = 0.055$; Fig. 2c).

To examine these results further and investigate the *diathesis* effect at different levels of stress, nominally significant GxE were plotted between PRS and categories of SLE (i.e, "none", "low", and "high" SLE reported; Fig. 3). Examining the interaction found in the full cohort (PRS at PGC-MDD GWAS p-threshold $= 0.01$), we detected a significant direct *diathesis* effect on the risk of depressive symptoms in those participants reporting SLE, with a higher risk when greater numbers of SLE were reported ("low" number of SLE reported: PRS' $\beta = 0.043$, s.e. $= 0.021$, $p = 0.039$; "high" number of SLE reported: PRS' $\beta = 0.142$, s.e. $= 0.039$, $p = 2.86 \times 10^{-4}$; see Table 1 and Fig. 3a). Whereas, in participants who reported no SLE over the preceding 6 months, the risk of depressive symptoms was the same regardless of their *diathesis* risk ("none" SLE reported: PRS' $\beta = 0.021$, s.e. $= 0.022$, $p = 0.339$). Stratifying these results by sex, we found the same pattern as in the full cohort in women ("none": $p = 0.687$; "low": $p = 0.023$; "high": $p = 2 \times 10^{-3}$), but not in men ("none": $p = 0.307$; "low": $p = 728$; "high": $p = 0.053$; see Table 1 and Fig. 3a). However, the lack of a significant *diathesis* effect in men may be due to their lower sample size and its corresponding reduced power.

Examining the interaction with PRS at PGC-MDD GWAS p-threshold $= 1 \times 10^{-5}$, with which a significant

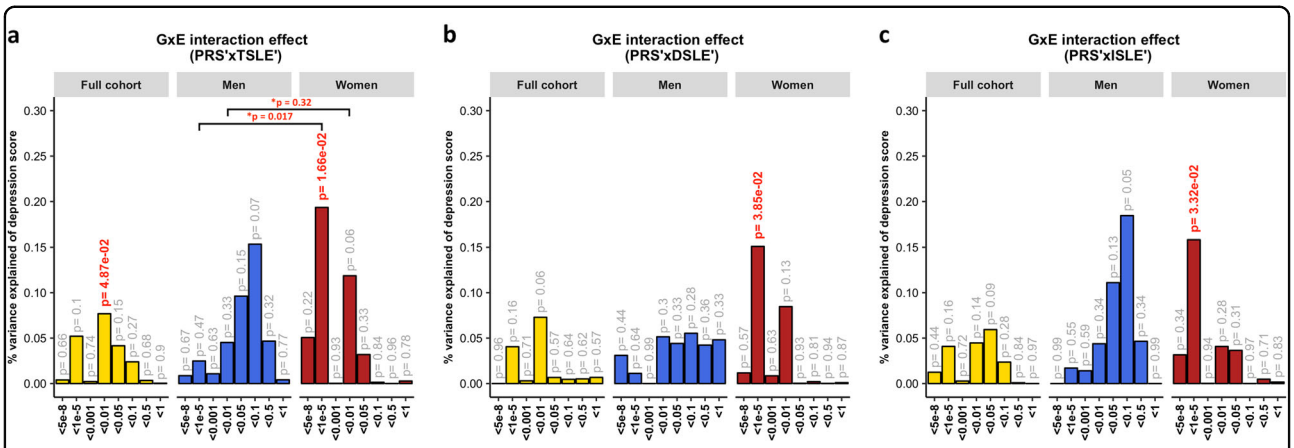


Fig. 2 Association between GxE effect and depression score. The plots show the percentage of depression score explained by the interaction term (two-sided tests) fitted in linear mixed models to empirically test the *diathesis-stress* model. Red numbers show significant interactions *p*-values. *Shows the significance of the difference in variance explained between sexes. The full cohort (yellow) was split into men (blue) and women (red). PRS were generated at 8 *p*-threshold levels using summary statistics from the Psychiatric Genetic Consortium MDD GWAS (released July 2016) with the exclusion of Generation Scotland participants. The interaction effect was tested with the number of: (**a**) SLE (TSLE), (**b**) "dependent" SLE (DSLE) and **c**) "independent" SLE (ISLE). In Colodro-Conde et al., the variance of depression score explained in their sample by GxE was 0.12% ($p = 7 \times 10^{-3}$). GxE were also significant in women ($p = 2 \times 10^{-3}$) explaining up to 0.25% of depression score variation, but not in men ($p = 0.059$; $R^2 = 0.17\%$; negative/protective effect on depression score)

interaction was detected in women, we detected a significant *diathesis* effect on depression score only when stratifying by sex in those participants who did not reported SLE over the last 6 months (see Table 1). The *diathesis* effect was positive in men (PRS' $\beta = 0.082$, s.e. = 0.034, $p = 0.016$, $R^2 = 0.7\%$; Fig. 3b), consistent with the contribution of risk alleles. Conversely, the *diathesis* effect was negative in women (PRS' β = -0.061, s.e. = 0.029, p = 0.037, $R^2 = 0.4\%$; Fig. 3b), suggesting a protective effect of increasing PRS in those women reporting no SLE, and consistent with the contribution of alleles to individual sensitivity to both positive and negative environmental effects (i.e., "plasticity alleles" rather than "risk alleles")[52,53]. This PRS accounted for the effect of just 34 SNPs, and the size of its GxE across sexes were significantly different (GxE*sex $p = 0.017$; Fig. 2a), supporting possible differences in the underlying stress-response mechanisms between women and men.

Discussion

The findings reported in this study support those from Colodro-Conde et al.[23], in an independent sample of similar sample size and study design, and also supports possible sex-specific differences in the effect of genetic risk of MDD in response to SLE.

Both Colodro-Conde et al. and our study suggest that individuals with an inherent genetic predisposition to MDD, reporting high number of recent SLE, are at additional risk of depressive symptoms due to GxE effects, thus validating the *diathesis-stress* theory. We identified nominally significant GxE effects in liability to depression at the population level ($p = 0.049$) and in women (p = 0.017), but not in men ($p = 0.072$). However, these interactions did not survive multiple testing correction ($p > 2.08 \times 10^{-3}$) and the power of both studies to draw robust conclusions remains limited[54]. With increased power these studies could determine more accurately both the presence and magnitude of a GxE effect in depression. To better understand the effect of PRS at different levels of exposure to stress, we examined the nominally significant interactions detected in the full sample by categorizing participants on three groups based on the number of SLE reported (i.e., "none", "low" or "high"). We detected a significant *diathesis* effect on risk of depression only in those participants reporting SLE, but not in those participants that reported no SLE over the preceding 6 months. Furthermore, the *diathesis* effect was stronger on those participants reporting a "high" number of SLE ($\beta = 0.142$, $p = 2.86 \times 10^{-4}$) compared to those participants reporting a "low" number of SLE ($\beta = 0.043$, $p = 0.039$). The former effect was significant and survived a conservative Bonferroni correction to adjust for multiple testing ($p < 6.94 \times 10^{-4}$). This finding corroborates the *diathesis-stress* model for depression and supports the results of Colodro-Conde et al. in an independent sample.

To investigate the relative contribution of the GxE to the variance of depression, we examined in the full cohort the total variance of depression score explained by the PRS main effect and the significant GxE effect jointly. Together, they explained 0.34% of the variance, of which 0.07% of the variance of the depression score was attributed to the GxE effect (p-threshold = 0.01; PRS $p = 1.19 \times 10^{-4}$, GxE p =

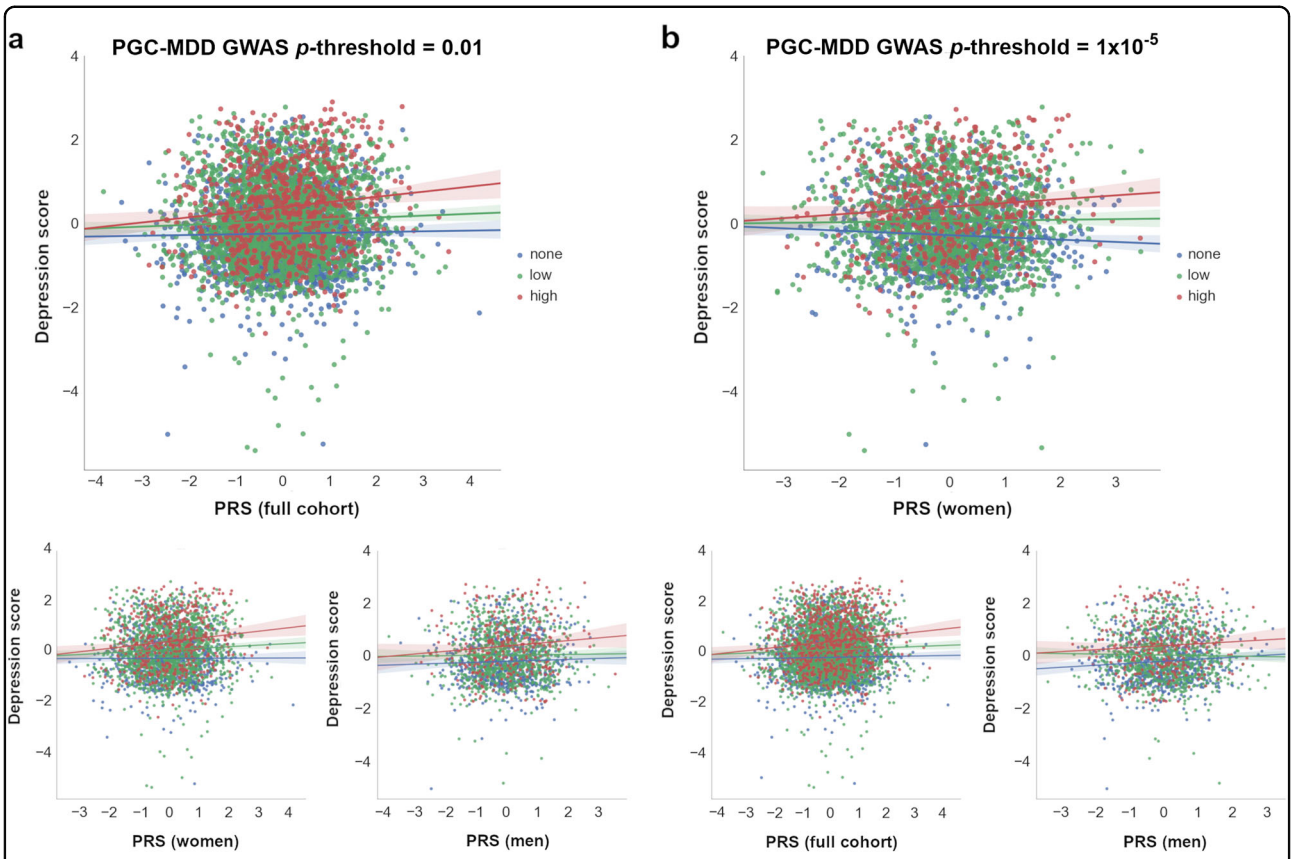


Fig. 3 Scatterplot of *diathesis-stress* interactions on depression score. Interactions with PRS at which nominally significant GxE effects were detected in (**a**) full cohort (p-threshold = 0.01) and (**b**) in women (p-threshold = 1×10^{-5}) are shown. At bottom, the remaining samples (i.e., *full cohort*, *women* or *men*) at same p-threshold are shown for comparison. The X-axis represents the direct effect of PRS (standard deviation from the mean) based on (**a**) p-threshold = 0.01 and (**b**) p-threshold = 1×10^{-5}, using the total number of SLE reported by each participant (dot) as environmental exposures at three SLE levels represented by colors. Blue: 0 SLE, "no stress", $n = 1\,833/1\,041/792$; green: 1 or 2 SLE, "low stress", $n = 2\,311/1\,459/852$; red: 3 or more SLE, "high stress", $n = 775/490/285$; in the full cohort, women and men, respectively. *Y*-axis reflects the depression score standardized to mean of 0 and standard deviation of 1. Lines represent the increment in risk of depression at a certain degree of "stress" dependent on a genetic predisposition (= *diathesis*)

0.049; both derived from the full diathesis-model with TSLE). This is lower than the proportion of variance attributed to common SNPs (8.9%) in the full PGC-MDD analysis[16]. As Colodro-Conde *et al.* noted, this result aligns with estimates from experimental organisms suggesting that around 20% of the heritability may be typically attributed to the effects of GxE[55], although it is inconsistent with twin studies of the majority of human traits with the potential exception of depression[56].

Consistent with PRS predicting "personal" SLE in Colodro-Conde et al., PRS for MDD predicted SLE in our study (see Supplementary Fig. 1), although not at the p-threshold at which significant GxE effects were detected. Genetic factors predisposing to MDD may contribute to individuals exposing themselves to, or showing an increased reporting of, SLE via behavioral or personality traits[57,58]. Such genetic mediation of the association between depression and SLE would disclose a gene-environment correlation (i.e., genetic effects on the probability of undergoing a SLE) that hinders interpretion of our findings as pure GxE effects[59,60]. To address this limitation and assess this aspect, following Colodro-Conde et al., we split the 12-items TSLE measure into SLE that are either potentially "dependent" of a participant's own behavior (DSLE; therefore, potentially driven by genetic factors) or not ("independent" SLE; ISLE)[45,48]. DSLE are reported to be more heritable and have stronger associations with MDD than ISLE[48,61,57]. In our sample, DSLE is significantly heritable ($h^2_{SNP} = 0.131$, s.e. = 0.071, $p = 0.029$), supporting a genetic mediation of the association, whereas ISLE is not significantly heritable ($h^2_{SNP} = 0.000$, s.e. = 0.072, $p = 0.5$)[62]. Nominally significant GxE effects were seen in women for both DSLE and ISLE, suggesting that both GxE and gene-environment correlation co-occur. Colodro-Conde et al. did not identify significant GxE using independent SLE as the exposure.

Table 1 *Diathesis* effect on depression score in three SLE categories

Sample	Full cohort[a]			Women			Men		
SLE category	**None**	**Low**	**High**	**None**	**Low**	**High**	**None**	**Low**	**High**
PRS at *p* value threshold = 0.01									
N	1833	*2311*	775	1041	*1459*	*490*	792	852	285
Effect	0.021	*0.043*	**0.142**	0.0118	*0.0617*	*0.1538*	0.0346	0.0113	0.1227
s.e.	0.022	*0.021*	**0.039**	0.029	*0.027*	*0.049*	0.034	0.032	0.063
t	0.957	*2.07*	**3.644**	0.403	*2.274*	*3.112*	1.021	0.348	1.947
p value	0.339	*0.039*	**2.86×10^{-4}**	0.687	*0.023*	*0.002*	0.307	0.728	0.053
CI (95%)	−0.022, 0.065	*0.002, 0.084*	**0.065, 0.218**	−0.046, 0.069	*0.008, 0.115*	*0.057, 0.251*	−0.032, 0.101	−0.052, 0.075	−0.001,0.247

Sample	Full cohort			Women[a]			Men		
SLE category	**None**	**Low**	**High**	**None**	**Low**	**High**	**None**	**Low**	**High**
PRS at *p* value threshold = 1×10^{-5}									
N	1833	2311	775	*1041*	1459	490	*792*	852	285
Effect	−0.0022	0.0032	0.0705	*−0.061*	0.014	0.078	*0.082*	−0.0176	0.0548
s.e.	0.022	0.021	0.04	*0.029*	0.027	0.049	*0.034*	0.033	0.07
t	−0.098	0.153	1.76	*−2.086*	0.541	1.609	*2.416*	−0.537	0.778
p value	0.922	0.878	0.079	*0.037*	0.589	0.108	*0.016*	0.592	0.437
CI (95%)	−0.046, 0.041	−0.037, 0.044	−0.008, 0.149	*−0.119, −0.004*	−0.038, 0.066	−0.017, 0.174	*0.015, 0.149*	−0.082,0.047	−0.084, 0.193

Note: Reported values at *p*-thresholds where nominally significant GxE effects were detected
[a]Sample where nominally significant GxE was detected. SLE categories (number of SLE reported): 0 SLE = "none", 1 or 2 SLE = "low", and 3 or more SLE = "high". In italic, nominally significant effects. In bold, robustly significant effect after conservative Bonferroni correction ($p < 6.94 \times 10^{-4}$)

Between-sex differences in stress response could help to explain previous differences seen between sexes in depression such as those in associated risk (i.e., approximately 1.5–2-fold higher in women), symptoms reported and/or coping strategies (e.g., whereas women tend to cope through verbal and emotional strategies, men tend to cope by doing sport and consuming alcohol)[63–67]. This also aligns with an increased risk associated with a lack of social support seen in women compared to men[23]. Furthermore, although we do not know whether participants experienced recent events with positive effects, we saw a protective effect in those women who did not experienced recent SLE ($p = 0.037$), suggesting that some genetic variants associated with MDD may operate as "plasticity alleles" and not just as "risk alleles"[52,53]. This effect was neutralized in the full cohort due to an opposite effect in men ($p = 0.016$), but it is supported by previous protective effects reported when using a serotoninergic multilocus profile score and absence of SLE in young women[68]. These findings would be consistent with a differential-susceptibility model of depression[69,70], also suggested by the interaction effects seen between the serotonin transporter linked promoter region gene (5-HTTLPR) locus and family support and liability to adolescent depression in boys[71]. However, our results and the examples given are only nominally significant and will require replication in larger samples. Robust identification of sex-specific differences in genetic stress-response could improve personalized treatments and therapies such as better coping strategies.

There are notable differences between our study and Colodro-Conde et al. to consider before accepting our findings as a replication of their results. First, differences in PRS profiling may have affected replication power. We used the same equivalent PGC-MDD2 GWAS as discovery sample. However, whereas Colodro-Conde et al. generated PRS in their target sample containing over 9.5 M imputed SNP, in this study we generated PRS in a target sample of over 560 K genotyped SNPs (see Supplementary table 1 for comparison). This potentially results in a less informative PRS in our study, with less predictive power, although the variance explained by our PRS was slightly larger (0.64% vs. 0.46%). The size of the discovery sample is key to constructing an accurate predictive PRS, but to exploit the greatest number of the available variants may be an asset[54]. Secondly, different screening tools were used to measure both current depressive symptoms and recent environmental stressors across the two studies. Both studies transformed their data, using item response theory or by log-transformation, to improve the data distribution.

However, neither study used depression scores that were normally distributed. The scale of the instruments used and their corresponding parameterization when testing for an interaction could have a direct effect on the size and significance of their interaction;[55,72] so findings from GxE must be taken with caution. Furthermore, although both screening methods have been validated and applied to detect depressive symptoms, different questions may cover and emphasize different features of the illness, which may result in different results. The same applies to the measurement of environmental stressors in the two studies. Both covering of a longer time-period and upweighting by "dependent" SLE items may explain the increased explanatory power of "personal" SLE (12.9%) in Colodro-Conde et al. to predict depression score compared to our "total" SLE measure (4.91%). Finally, the unmeasured aspects of the exposure to SLE or its impact may also contribute to the lack of a stronger replication and positive findings.

In conclusion, despite differences in the measures used across studies, we saw concordance and similar patterns between our results and those of Colodro-Conde et al.[23] Our findings, therefore, add validity to the *diathesis-stress* theory for depression. Empirically demonstrating the *diathesis-stress* theory for depression would validate recent[20–22] and future studies using a genome-wide approach to identify genetic mechanisms and interactive pathways involved in GxE underpinning the causative effect of "stress" in the development of depressive symptoms and in mental illness in general. This study adds to our understanding of gene-by-environment interactions, although larger samples will be required to confirm differences in *diathesis-stress* effects between women and men.

Acknowledgements

A.A.S. is funded by the University of Edinburgh (www.ed.ac.uk) and MRC for his PhD study at the University of Edinburgh Institute of Genetics and Molecular Medicine (www.ed.ac.uk/igmm). D.J.M. acknowledges the financial support of NHS Research Scotland (NRS) through NHS Lothian. MA is supported by STRADL through a Wellcome Trust Strategic Award (reference 104036/Z/14/Z). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. The genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). The Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium depends on the contributions of many parties.

Author details

[1]Medical Genetics Section, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [2]Division of Psychiatry, Deanery of Clinical Sciences, Royal Edinburgh Hospital, University of Edinburgh, Morningside Park, Edinburgh EH10 5HF, UK. [3]Health Informatics Centre, University of Dundee, Dundee, UK. [4]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [5]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-018-0356-7).

Received: 5 April 2018 Revised: 28 November 2018 Accepted: 10 December 2018
Published online: 18 January 2019

References

1. Hammen, C. Stress and depression. *Annu Rev. Clin. Psychol.* **1**, 293–319 (2005).
2. Kessler, R. C. The effects of stressful life events on depression. *Annu Rev. Psychol.* **48**, 191–214 (1997).
3. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* **156**, 837–841 (1999).
4. Paykel, E. S. Life events and affective disorders. *Acta Psychiatr. Scand* **108**, 61–66 (2003).
5. Stroud, C. B., Davila, J. & Moyer, A. The relationship between stress and depression in first onsets versus recurrences: a meta-analytic review. *J. Abnorm. Psychol.* **117**, 206–213 (2008).
6. Ensel, W. M., Peek, M. K., Lin, N. & Lai, G. Stress in the life course: a life history approach. *J. Aging Health* **8**, 389–416 (1996).
7. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. *J. Nerv. Ment. Dis.* **186**, 661–669 (1998).
8. Mazure, C. M. Life stressors as risk factors in depression. *Clin. Psychol.* **5**, 291–313 (1998).
9. Lichtenberg, P. & Belmaker, R. H. Subtyping major depressive disorder. *Psychother. Psychosom.* **79**, 131–135 (2010).
10. Elisei, S., Sciarma, T., Verdolini, N. & Anastasi, S. Resilience and depressive disorders. *Psychiatr. Danub.* **25**(Suppl 2), S263–S267 (2013).
11. Monroe, S. M. & Simons, A. D. Diathesis-stress theories in the context of life stress research: implications for the depressive disorders. *Psychol. Bull.* **110**, 406–425 (1991).
12. Vogel, F. Schizophrenia genesis: the origins of madness. *Am. J. Human. Genet.* **48**, 1218–1218 (1991).
13. Mann, J. J., Waternaux, C., Haas, G. L. & Malone, K. M. Toward a clinical model of suicidal behavior in psychiatric patients. *Am. J. Psychiatry* **156**, 181–189 (1999).
14. Riemann, D. et al. The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep. Med Rev.* **14**, 19–31 (2010).
15. Bolt, M. A., Helming, L. M. & Tintle, N. L. The associations between self-reported exposure to the chernobyl nuclear disaster zone and mental health disorders in Ukraine. *Front. Psychiatry* **9**, 32 (2018).
16. Wray, N. R. & Sullivan, P. F. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668–681 (2018).
17. Peyrot, W. J. et al. Effect of polygenic risk scores on depression in childhood trauma. *Br. J. Psychiatry* **205**, 113–119 (2014).
18. Musliner, K. L. et al. Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychol. Med.* **45**, 1709–1720 (2015).
19. Peyrot, W. J. et al. Does childhood trauma moderate polygenic risk for depression? A Meta-analysis of 5765 Subjects From the Psychiatric Genomics Consortium. *Biol. Psychiatry* **84**, 138–147 (2018).
20. Dunn, E. C. et al. Genome-Wide Association Study (GWAS) and Genome-Wide by Environment Interaction Study (GWEIS) of Depressive Symptoms in African American and Hispanic/Latina Women. *Depress. Anxiety* **33**, 265–280 (2016).
21. Otowa, T. et al. The first pilot genome-wide gene-environment study of depression in the Japanese Population. *PLoS One* **11**, e0160823 (2016).
22. Ikeda, M. et al. Genome-wide environment interaction between depressive state and stressful life events. *J. Clin. Psychiatry* **77**, e29–e30 (2016).

23. Colodro-Conde, L. et al. A direct test of the diathesis-stress model for depression. *Mol. Psychiatry* **23**, 1590–1596 (2018).
24. Mullins, N. et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol. Med* **46**, 759–770 (2016).
25. Iyegbe, C., Campbell, D., Butler, A., Ajnakina, O. & Sham, P. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc. Psychiatry Psychiatr. Epidemiol.* **49**, 169–182 (2014).
26. McGrath, J. J., Mortensen, P. B., Visscher, P. M. & Wray, N. R. Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. *Schizophr. Bull.* **39**, 955–959 (2013).
27. Plomin, R. Commentary: missing heritability, polygenic scores, and gene-environment correlation. *J. Child Psychol. Psychiatry* **54**, 1147–1149 (2013).
28. Wray, N. R. et al. Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).
29. Smith, B. H. et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J. Epidemiol.* **42**, 689–700 (2013).
30. Gunderson, K. L. Whole-genome genotyping on bead arrays. In *DNA Microarrays for Biomedical Research: Methods and Protocols* (ed. Dufva, M.) 197–213 (Humana Press, Totowa, 2009).
31. Kerr, S. M. et al. Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med. Genet.* **14**, 38 (2013).
32. Nagy, R. et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* **9**, 23 (2017).
33. Navrady, L. B. et al. Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *Int. J. Epidemiol* **47**, 13–14g (2018).
34. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
35. Smith, B. H. et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* **7**, 74 (2006).
36. Fernandez-Pujals, A. M. et al. Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
37. Amador, C. et al. Recent genomic heritage in Scotland. *BMC Genom.* **16**, 437 (2015).
38. Goldberg, D. P. & Hillier, V. F. A scaled version of the General Health Questionnaire. *Psychol. Med.* **9**, 139–145 (1979).
39. Sterling, M. General Health Questionnaire - 28 (GHQ-28). *J. Physiother.* **57**, 259 (2011).
40. Goldberg, D. P. et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol. Med.* **27**, 191–197 (1997).
41. Banks, M. H. Validation of the General Health Questionnaire in a young community sample. *Psychol. Med.* **13**, 349–353 (1983).
42. Marks, A. D. G., Horrocks, K. A. & Schutte, N. S. Emotional intelligence mediates the relationship between insecure attachment and subjective health outcomes. *Personal. Individ. Differ.* **98**, 188–192 (2016).
43. O'Rourke, S., MacHale, S., Signorini, D. & Dennis, M. Detecting psychiatric morbidity after stroke: comparison of the GHQ and the HAD Scale. *Stroke* **29**, 980–985 (1998).
44. Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B. & Wittchen, H.-U. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *Int. J. Methods Psychiatr. Res.* **7**, 171–185 (1998).
45. Brugha, T., Bebbington, P., Tennant, C. & Hurry, J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol. Med.* **15**, 189–194 (1985).
46. Brugha, T. S. & Cragg, D. The list of threatening experiences: the reliability and validity of a brief life events questionnaire. *Acta Psychiatr. Scand.* **82**, 77–81 (1990).
47. Motrico, E. et al. Psychometric properties of the List of Threatening Experiences--LTE and its association with psychosocial factors and mental disorders according to different scoring methods. *J. Affect Disord.* **150**, 931–940 (2013).
48. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. The assessment of dependence in the study of stressful life events: validation using a twin design. *Psychol. Med.* **29**, 1455–1460 (1999).
49. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
50. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet* **88**, 76–82 (2011).
51. Keller, M. C. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24 (2014).
52. Belsky, J. & Beaver, K. M. Cumulative-genetic plasticity, parenting and adolescent self-regulation. *J. Child Psychol. Psychiatry* **52**, 619–626 (2011).
53. Belsky, J. et al. Vulnerability genes or plasticity genes? *Mol. Psychiatry* **14**, 746–754 (2009).
54. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**, e1003348 (2013).
55. Eaves, L. J., Last, K., Martin, N. G. & Jinks, J. L. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Br. J. Math. Stat. Psychol.* **30**, 1–42 (1977).
56. Polderman, T. J. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
57. Clarke, T. et al. Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism [version 2; referees: 3 approved with reservations]. *Wellcome Open Res* **3**, 11 (2019).
58. Kendler, K. S., Kuhn, J. & Prescott, C. A. The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *Am. J. Psychiatry* **161**, 631–636 (2004).
59. Kendler, K. S. & Eaves, L. J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am. J. Psychiatry* **143**, 279–289 (1986).
60. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* **84**, 309–322 (1977).
61. Plomin, R., Lichtenstein, P., Pedersen, N. L., McClearn, G. E. & Nesselroade, J. R. Genetic influence on life events during the last half of the life span. *Psychol. Aging* **5**, 25–30 (1990).
62. Arnau Soler, A. et al. Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Transl. Psychiatry* (in press).
63. Weissman, M. M. et al. Sex differences in rates of depression: cross-national perspectives. *J. Affect Disord.* **29**, 77–84 (1993).
64. Van de Velde, S., Bracke, P. & Levecque, K. Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Soc. Sci. Med.* **71**, 305–313 (2010).
65. Labonte, B. et al. Sex-specific transcriptional signatures in human depression. *Nat. Med* **24**, 525 (2018).
66. Angst, J. et al. Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur. Arch. Psychiatry Clin. Neurosci.* **252**, 201–209 (2002).
67. Piccinelli, M. & Wilkinson, G. Gender differences in depression. *Crit. Rev. Br. J. Psychiatry* **177**, 486–492 (2000).
68. Vrshek-Schallhorn, S. et al. Additive genetic risk from five serotonin system polymorphisms interacts with interpersonal stress to predict depression. *J. Abnorm. Psychol.* **124**, 776–790 (2015).
69. Belsky, J. & Pluess, M. Beyond diathesis stress: differential susceptibility to environmental influences. *Psychol. Bull.* **135**, 885–908 (2009).
70. Belsky, J., Bakermans-Kranenburg, M. J. & van Ijzendoorn, M. H. For better and for worse: differential susceptibility to environmental influences. *Curr. Dir. Psychol. Sci.* **16**, 300–304 (2007).
71. Li, J. J., Berk, M. S. & Lee, S. S. Differential susceptibility in longitudinal models of gene-environment interaction for adolescent depression. *Dev. Psychopathol.* **25**, 991–1003 (2013).
72. Kang, S.-M. & Waller, N. G. Moderated multiple regression, spurious interaction effects, and IRT. *Applied Psychological Measurement* **29**, 87–105 (2005).

Appendix D

Appendix D contains supplementary information and supplementary figures for **chapter 5**: Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. The published article in *Translational Psychiatry* is also included.

D.1 R code to perform GWEIS.

The R script I used to perform GWEIS is a modified version from the original R script developed by Almli *et al.*[1] (https://epstein-software.github.io/robust-joint-interaction). The modifications added into the code output beta coefficients from both SNP and SNPxE terms and the *p*-value for just the interaction (GxE) term into the final GWEIS summary statistics.

Using the original R script (version 20 Apr 2017), a version of the script as the one we used can be achieved by replacing from the source code:

r <- c(p_model,p_robust)

with:

wald_int_only<-((gee.fit$beta[3])^2)/gee.fit$vbeta[3,3] # this is the robust wald test for just the interaction parameter only

p_int_only<-pchisq(wald_int_only,1,lower.tail=F) # this is the p-value for the wald test

r <- c(p_model,p_robust,beta_est[c(2,3)],p_int_only) # The 'beta_est' vector contains all the regression coefficients for all parameters in the GEE model.

These additional lines of code were courtesy of Prof. Michael Epstein[1].

Reference

1. Almli, L.M. *et al.* Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392-9 (2014).

D.2 Statistical analyses

Post-GWAS/GWEIS analyses. GWAS and GWEIS (for both GxE and joint effects) summary statistics were analysed using FUMA[1] (http://fuma.ctglab.nl). Gene-based tests were conducted using MAGMA[2] through FUMA platform using the default MAGMA settings. Genome-wide significance at gene-based test was set at the Bonferroni-corrected significance threshold $p = 0.05/18{,}068 = 2.77 \times 10^{-6}$. FUMA was also used to assess functional annotation, gene prioritization and pathway enrichment using lead SNPs from associated genomic loci. SNPs were clumped according to linkage disequilibrium ($r^2 = 0.6$) using default settings to identify independent lead SNPs with a $p < 1 \times 10^{-5}$. Gene mapping was performed in protein-coding genes under default settings using: positional mapping, eQTL mapping in 10 brain tissues from GTEx, and chromatin interaction mapping using Hi-C data from two brain regions: dorsolateral/prefrontal cortex and hippocampus. The MHC region (25-34 Mb) was excluded from these analyses. We assessed gene-set enrichment of: differentially expressed genes in 53 tissue types, Canonical Pathways, and Gene Ontology terms. Previously reported associations from the GWAS catalog[3] were reported for independent lead SNPs and/or their proxy SNPs.

Polygenic profiling & prediction. Polygenic risk scores (PRS) were constructed using PRSice-2[4] (clump-based pruning $r^2 = 0.1$, 10MB per window) for thirteen *p* thresholds (<0.001, <0.005, <0.01, <0.02, <0.03, <0.04, <0.05, <0.1, <0.2, <0.3, <0.4, <0.5, <=1) and standardized to a mean of 0. We calculated the proportion of variance explained using R^2 and Nagelkerke's R^2

coefficients for binary traits. Polygenic risk scores weighting by GxE effects (PRS_{GxE}) were generated as standard PRS[5] using summary statistics from GWEIS as input (discovery sample), instead of summary statistics from GWAS. Thus, weighting PRS by GxE effects (or joint effects) instead of main additive effects.

References

1. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
2. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
3. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
4. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-8 (2015).
5. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**, e1003348 (2013).

D.3 Evidence supporting a link between stress and depression in stress-related phenotypes

We hypothesized that pleiotropy between stress-related conditions, both mental and physical, may be due in part to shared genetic stress-response mechanisms. In this study we report evidence, albeit weak, for an effect of GxE involved in depressive symptoms into other stress-related traits, in particular for schizotypal personality, heart disease and chronic obstructive pulmonary disease (COPD). We believe that at least partially, genetics underlying GxE detected in GWEIS of depressive symptoms belongs to unique stress-response mechanisms also involved in the aetiology of these other phenotypes. There are many studies on the literature that provide evidence supporting a link between stress and depression in these phenotypes that supports our conclusion.

Schizotypal personality. One of the main characteristics of schizotypal personality is an abnormal perception of experiences and dysfunctional coping strategies[6-8], which overlaps with autistic personality traits, anxiety and depression[9-11]. Schizotypal personality correlates with SLE and negative coping strategies in healthy individuals[12]. The effect of stress on schizotypal personality traits is moderated by genetics[13] and mediated by dysfunctional coping[12,13]. Individuals with high schizotypal personality have been shown to report increased subjective stress and to display blunted and delayed cortisol response after acute psychosocial stress, compared to individuals with low schizotypal personality, and have higher baseline cortisol levels, suggesting a limited physiologically adaptation to SLE and a feasible

link with cardiovascular comorbidities through chronic overactivation of the sympathetic nervous system[14].

Heart disease. A direct relationship between the effects of psychological stress and cardiovascular disease has been extensively reported[15-21]. Depression, as well as anger and anxiety, may be a manifestation form underlying negative emotions, and it has been proposed that such underlying negativity may origin the link between these conditions and cardiovascular disease[22]. Psychosocial stress associated with anxiety disorders increase autonomic arousal via the hypothalamic-pituitary axis and circulating cathecolamines, which is associated to an elevated risk of pro-inflammatory state and hypertension leading to the development of coronary heart disease[23]. Previous studies also link anxiety disorder to cardiovascular disease[16], and evidence suggests a potential role for inflammation as an underlying mechanism linking chronic stress and an associated increased risk of coronary heart disease[16,17,24]. Psychological stress is a risk factor of hypertension[18,25], acute myocardial infarction[21], pathogenesis of coronary artery disease[16], cardiovascular morbidity and mortality[19], within other cardiovascular disease, with a potential elevated risk in women[20]. A consistent relationship between the occurrence of major depression episodes and cardiovascular disease has been demonstrated, and evidence suggest that there is a continuum spectrum of risk for coronary artery disease associated with depression according to the magnitude of depressive symptoms[16]. Smoking is a coping strategy[26] and behavioural risk factor associated with depression[27], and risk of suffering a cardiac event is substantially higher in depressed individuals who smoke than in depressed individuals who do not smoke[28].

Chronic obstructive pulmonary disease (COPD). In this study, we suggested that GxE effect in depressive symptoms is related with risk of COPD. The prevalence of depressive symptoms, depression and risk of anxiety is higher among COPD patients[29,30,31]. SLE have been associated with increased depressive symptoms, anxiety and worse quality of life in individuals with COPD compared to healthy controls[32,33], with significant SLE-by-COPD interaction suggesting a substantially greater detrimental effect of stress in those individuals suffering COPD[32]. Participants with COPD have reported to experience the same number and severity of non-illness-related occurrence of SLE and perceived stress as non-COPD healthy participants, suggesting that individuals with COPD may be more vulnerable to detrimental effects of SLE than healthy controls[32]. The experience of anxiety and distress produced by breathing difficulties is an intrinsic source of stress directly related to COPD[34,35]. However, poor coping skills may be the principal psychological problem that contributes to psychological distress and poor quality of life among COPD patients[36]. Depression and heart failure are associated with disease-specific health-related quality of life in COPD patients[37]. Smoking is recognized as the most important risk factor for COPD[38]. Lifetime cumulative stress, including psychological stress early in life, has been shown to have and additive effect over time on risk for heavy smoking[39]. Elevated risk for developing nicotine dependence has been shown in individuals who experienced trauma, and exacerbated by post-traumatic stress disorder[40]. Cigarette smoke disrupts homeostasis of the alveolar capillary unit and generates endogenous mediators of inflammation that could result in COPD[41]. The exposure to inhaled toxics from smoke activates a cellular integrated stress response implicated in pulmonary pathology and lung disease[42]. However, well-known

behavioural risk factors such as smoking become extremely difficult to quit in those individuals with high life stress and depression levels[27,43]. In addition, although smokers often report smoking as a strategy to relieve stress, nicotine dependency is reported to exacerbate psychological stress, showing stress levels of smokers being higher than those of non-smokers and reporting worse mood, specially between cigarettes[44].

Interestingly, a post hoc test conducted in our study showed that PRS weighted by GxE effect derived from Generation Scotland GWEIS using TSLE that predicts COPD also predicts number of cigarettes smoked in an average week in 75 smokers from an independent GS sample (p = 4.43x10^{-6}). This was not replicated using the UKB GxE effect. However, it suggests that genetic stress-response and stress-related traits may be partially related through GxE driven coping styles. Increased risk for smoking is also linked to other psychiatric disorders where smoking may play an important role either as a causal factor, as an agent promoting brain changes, or as modulator of medication's effect[45]. Asthma, a separate respiratory disease, has also shown robust relationships with psychological stress, depression, neuro-psychiatric factors and neuroendocrine, immunologic and inflammatory systems[24,46-50].

***Other stress-related traits*.** Chronic psychological stress has been recognised to influence the immune system and to have an impact on many inflammatory process and disease, and dysregulated inflammatory processes have been linked to major depressive disorder and its comorbidities (e.g. asthma, diabetes, cardiovascular disorders, stroke)[24,51]. Therefore, the list of stress-related conditions link to depression through genetic stress-response mechanisms may be larger, including

other personality traits such as extraversion or some cancers, among others. Findings suggest that extraversion was improved by incorporating SLE GxE effects into main additive effect as a combined joint effect; prediction otherwise missed by any GxE effect or UKB main (depressive symptoms) additive effects alone. Extraversion mediates (both positive and negative) life events and resilience, low extraversion being associated with higher numbers of SLE[24]. High levels of extraversion, which involves energy, positive emotionality and dominance, is linked with stress and adversity coping strategies through positive affective style, and related to resilience and higher level of social interaction allowing an environment of positive networks of social support[52]. Indeed, individuals who are genetically predisposed to high level of extraversion my perceive (and report) life events as more controllable and positive; which in turn may lead to higher level of extraversion[53]. Low extraversion is associated with depression and depressive symptoms[54,55]. Another interesting case is the link between stress and cancer. Stress has been associated with cancer[56-59]. A recent study with lifetime exposure has reported that perceived stress, principally from workplace psychological stress over cumulative duration of more than 15 years, was significantly positively associated with greater odds of cancer (including lung, colon, rectum, stomach cancers and Non-Hodgkin lymphoma) among men[57]. Interesting, piRNAs and PIWI-like proteins, including HIWI2 encoded by *PIWIL4,* a gene involved in chromatin-modification[60] and located at the locus with the strongest genome-wide association between GxE effect and depressive symptoms reported in this study has been suggested as biomarkers and potential therapeutic targets for breast cancer[61,62] and gastric cancer[63]. The effects of stress may have a pronounced impact on epigenetic changes like

chromatin or histone modifications. Epigenetic processes have been widely implicated in cancer[64,65].

In summary. As reviewed, psychological stress is recognized to play a role in the risk and prognosis of many chronic illnesses, as well as depressive symptoms, including evidence for other stress-related comorbidities of depression not detailed above such as diabetes[66-71] and stroke[72-74], among others[75]. Furthermore, stress may influence the extent to which individuals engage with behavioural risk factors as coping strategies; including smoking, fatty diet intake, excessive alcohol intake or substance abuse. GxE effects and genetic stress-perception, stress-exposure or stress-response may drive selection of these coping strategies and modulate coping skills and how we work through SLE, and negative emotions.

D.4 References

1. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
2. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
3. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
4. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-8 (2015).
5. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**, e1003348 (2013).
6. Bijttebier, P. & Vertommen, H. Coping strategies in relation to personality disorders. *Personality and Individual Differences* **26**, 847-856 (1999).
7. Laloyaux, J., Dessart, G., Van der Linden, M., Lemaire, M. & Laroi, F. Maladaptive emotion regulation strategies and stress sensitivity mediate the relation between adverse life events and attenuated positive psychotic symptoms. *Cogn Neuropsychiatry* **21**, 116-29 (2016).
8. MacAulay, R. & Cohen, A.S. Affecting coping: does neurocognition predict approach and avoidant coping strategies within schizophrenia spectrum disorders? *Psychiatry Res* **209**, 136-41 (2013).
9. Mealey, A., Abbott, G., Byrne, L.K. & McGillivray, J. Overlap between autistic and schizotypal personality traits is not accounted for by anxiety and depression. *Psychiatry Res* **219**, 380-5 (2014).
10. Hofvander, B. *et al.* Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC Psychiatry* **9**, 35 (2009).
11. Buckley, P.F., Miller, B.J., Lehrer, D.S. & Castle, D.J. Psychiatric comorbidities and schizophrenia. *Schizophr Bull* **35**, 383-402 (2009).
12. Armando, M. *et al.* Coping Strategies Mediate the Effect of Stressful Life Events on Schizotypal Traits and Psychotic Symptoms in 22q11.2 Deletion Syndrome. *Schizophr Bull* (2018).
13. Hatzimanolis, A. *et al.* Stress-Dependent Association Between Polygenic Risk for Schizophrenia and Schizotypal Traits in Young Army Recruits. *Schizophr Bull* **44**, 338-347 (2018).
14. Walter, E.E., Fernandez, F., Snelling, M. & Barkus, E. Stress induced cortisol release and schizotypy. *Psychoneuroendocrinology* **89**, 209-215 (2018).
15. Dimsdale, J.E. Psychological stress and cardiovascular disease. *J Am Coll Cardiol* **51**, 1237-46 (2008).
16. Rozanski, A., Blumenthal, J.A. & Kaplan, J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation* **99**, 2192-217 (1999).

17. Wirtz, P.H. & von Kanel, R. Psychological Stress, Inflammation, and Coronary Heart Disease. *Curr Cardiol Rep* **19**, 111 (2017).
18. Imumorin, I.G. *et al.* A gene-environment interaction model of stress-induced hypertension. *Cardiovasc Toxicol* **5**, 109-32 (2005).
19. Brunner, E.J. Social factors and cardiovascular morbidity. *Neurosci Biobehav Rev* **74**, 260-268 (2017).
20. Varghese, T., Hayek, S.S., Shekiladze, N., Schultz, W.M. & Wenger, N.K. Psychosocial Risk Factors Related to Ischemic Heart Disease in Women. *Curr Pharm Des* **22**, 3853-70 (2016).
21. Rosengren, A. *et al.* Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 953-62 (2004).
22. Suls, J. & Bunde, J. Anger, anxiety, and depression as risk factors for cardiovascular disease: the problems and implications of overlapping affective dispositions. *Psychol Bull* **131**, 260-300 (2005).
23. Player, M.S. & Peterson, L.E. Anxiety disorders, hypertension, and cardiovascular risk: a review. *Int J Psychiatry Med* **41**, 365-77 (2011).
24. Straub, R.H. & Cutolo, M. Psychoneuroimmunology-developments in stress research. *Wien Med Wochenschr* **168**, 76-84 (2018).
25. Babu, G.R. *et al.* Is hypertension associated with job strain? A meta-analysis of observational studies. *Postgrad Med J* **90**, 402-9 (2014).
26. Long, D. Smoking as a coping strategy. *Nurs Times* **99**, 50, 53 (2003).
27. Glassman, A.H. *et al.* Smoking, smoking cessation, and major depression. *JAMA* **264**, 1546-9 (1990).
28. Anda, R. *et al.* Depressed affect, hopelessness, and the risk of ischemic heart disease in a cohort of U.S. adults. *Epidemiology* **4**, 285-94 (1993).
29. Zhang, M.W., Ho, R.C., Cheung, M.W., Fu, E. & Mak, A. Prevalence of depressive symptoms in patients with chronic obstructive pulmonary disease: a systematic review, meta-analysis and meta-regression. *Gen Hosp Psychiatry* **33**, 217-23 (2011).
30. van Ede, L., Yzermans, C.J. & Brouwer, H.J. Prevalence of depression in patients with chronic obstructive pulmonary disease: a systematic review. *Thorax* **54**, 688-92 (1999).
31. Eisner, M.D. *et al.* Influence of anxiety on health outcomes in COPD. *Thorax* **65**, 229-34 (2010).
32. Lu, Y. *et al.* Life event stress and chronic obstructive pulmonary disease (COPD): associations with mental well-being and quality of life in a population-based study. *BMJ Open* **2**(2012).
33. Yu, T., Frei, A., Ter Riet, G. & Puhan, M.A. Impact of Stressful Life Events on Patients with Chronic Obstructive Pulmonary Disease. *Respiration* **95**, 73-79 (2018).
34. DeVito, A.J. Dyspnea during hospitalizations for acute phase of illness as recalled by patients with chronic obstructive pulmonary disease. *Heart Lung* **19**, 186-91 (1990).
35. Gurney-Smith, B., Cooper, M.J. & Wallace, L.M. Anxiety and Panic in Chronic Obstructive Pulmonary Disease: The Role of Catastrophic Thoughts. *Cognitive Therapy and Research* **26**, 143-155 (2002).

36. Andenaes, R., Kalfoss, M.H. & Wahl, A.K. Coping and psychological distress in hospitalized patients with chronic obstructive pulmonary disease. *Heart Lung* **35**, 46-57 (2006).
37. Urff, M., van den Berg, J.W., Uil, S.M., Chavannes, N.H. & Damoiseaux, R.A. Depression and heart failure associated with clinical COPD questionnaire outcome in primary care COPD patients: a cross-sectional study. *NPJ Prim Care Respir Med* **24**, 14066 (2014).
38. Laniado-Laborin, R. Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21 century. *Int J Environ Res Public Health* **6**, 209-24 (2009).
39. Lloyd, D.A. & Taylor, J. Lifetime cumulative adversity, mental health and the risk of becoming a smoker. *Health (London)* **10**, 95-112 (2006).
40. Breslau, N., Davis, G.C. & Schultz, L.R. Posttraumatic stress disorder and the incidence of nicotine, alcohol, and other drug disorders in persons who have experienced trauma. *Archives of General Psychiatry* **60**, 289-294 (2003).
41. Tuder, R.M. & Yoshida, T. Stress responses affecting homeostasis of the alveolar capillary unit. *Proc Am Thorac Soc* **8**, 485-91 (2011).
42. van 't Wout, E.F., Hiemstra, P.S. & Marciniak, S.J. The integrated stress response in lung disease. *Am J Respir Cell Mol Biol* **50**, 1005-9 (2014).
43. Siahpush, M. & Carlin, J.B. Financial stress, smoking cessation and relapse: results from a prospective study of an Australian national sample. *Addiction* **101**, 121-7 (2006).
44. Parrott, A.C. Does cigarette smoking cause stress? *Am Psychol* **54**, 817-20 (1999).
45. Boksa, P. Smoking, psychiatric illness and the brain. *J Psychiatry Neurosci* **42**, 147-149 (2017).
46. Plourde, A., Lavoie, K.L., Raddatz, C. & Bacon, S.L. Effects of acute psychological stress induced in laboratory on physiological responses in asthma populations: A systematic review. *Respir Med* **127**, 21-32 (2017).
47. Sandberg, S., Jarvenpaa, S., Penttinen, A., Paton, J.Y. & McCann, D.C. Asthma exacerbations in children immediately following stressful life events: a Cox's hierarchical regression. *Thorax* **59**, 1046-51 (2004).
48. Wang, T. *et al.* Transcriptomic profiling of peripheral blood CD4(+) T-cells in asthmatics with and without depression. *Gene* **565**, 282-7 (2015).
49. Chida, Y., Hamer, M. & Steptoe, A. A bidirectional relationship between psychosocial factors and atopic disorders: a systematic review and meta-analysis. *Psychosom Med* **70**, 102-16 (2008).
50. Miyasaka, T., Dobashi-Okuyama, K., Takahashi, T., Takayanagi, M. & Ohno, I. The interplay between neuroendocrine activity and psychological stress-induced exacerbation of allergic asthma. *Allergol Int* **67**, 32-42 (2018).
51. Penninx, B.W., Milaneschi, Y., Lamers, F. & Vogelzangs, N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med* **11**, 129 (2013).

52. Campbell-Sills, L., Cohan, S.L. & Stein, M.B. Relationship of resilience to personality, coping, and psychiatric symptoms in young adults. *Behav Res Ther* **44**, 585-99 (2006).
53. Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A. & Spinath, F.M. Life events as environmental States and genetic traits and the role of personality: a longitudinal twin study. *Behav Genet* **42**, 57-72 (2012).
54. Grav, S., Stordal, E., Romild, U.K. & Hellzen, O. The relationship among neuroticism, extraversion, and depression in the HUNT Study: in relation to age and gender. *Issues Ment Health Nurs* **33**, 777-85 (2012).
55. Jylha, P. & Isometsa, E. The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depress Anxiety* **23**, 281-9 (2006).
56. Shin, K.J. *et al.* Molecular Mechanisms Underlying Psychological Stress and Cancer. *Curr Pharm Des* **22**, 2389-402 (2016).
57. Blanc-Lapierre, A. *et al.* Lifetime report of perceived stress at work and cancer among men: A case-control study in Montreal, Canada. *Prev Med* **96**, 28-35 (2017).
58. Blanc-Lapierre, A., Rousseau, M.C. & Parent, M.E. Perceived Workplace Stress Is Associated with an Increased Risk of Prostate Cancer before Age 65. *Front Oncol* **7**, 269 (2017).
59. Powell, N.D., Tarr, A.J. & Sheridan, J.F. Psychosocial stress and inflammation in cancer. *Brain Behav Immun* **30 Suppl**, S41-7 (2013).
60. Sugimoto, K. *et al.* The induction of H3K9 methylation by PIWIL4 at the p16Ink4a locus. *Biochem Biophys Res Commun* **359**, 497-502 (2007).
61. Krishnan, P. *et al.* Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. *Oncotarget* **7**, 37944-37956 (2016).
62. Wang, Z., Liu, N., Shi, S., Liu, S. & Lin, H. The Role of PIWIL4, an Argonaute Family Protein, in Breast Cancer. *J Biol Chem* **291**, 10646-58 (2016).
63. Wang, Y. *et al.* The PIWI protein acts as a predictive marker for human gastric cancer. *Int J Clin Exp Pathol* **5**, 315-25 (2012).
64. Sharma, S., Kelly, T.K. & Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).
65. Wei, J. *et al.* Discovery and Validation of Hypermethylated Markers for Colorectal Cancer. *Dis Markers* **2016**, 2192853 (2016).
66. Hackett, R.A. & Steptoe, A. Type 2 diabetes mellitus and psychological stress - a modifiable risk factor. *Nat Rev Endocrinol* **13**, 547-560 (2017).
67. Kelly, S.J. & Ismail, M. Stress and type 2 diabetes: a review of how stress contributes to the development of type 2 diabetes. *Annu Rev Public Health* **36**, 441-62 (2015).
68. Bergmann, N., Gyntelberg, F. & Faber, J. The appraisal of chronic stress and the development of the metabolic syndrome: a systematic review of prospective cohort studies. *Endocr Connect* **3**, R55-80 (2014).
69. Pan, A. *et al.* Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies. *Diabetes Care* **35**, 1171-80 (2012).

70. Walker, R.J., Strom Williams, J. & Egede, L.E. Influence of Race, Ethnicity and Social Determinants of Health on Diabetes Outcomes. *Am J Med Sci* **351**, 366-73 (2016).
71. Nyberg, S.T. *et al.* Job strain as a risk factor for type 2 diabetes: a pooled analysis of 124,808 men and women. *Diabetes Care* **37**, 2268-75 (2014).
72. Surtees, P.G. *et al.* Psychological distress, major depressive disorder, and risk of stroke. *Neurology* **70**, 788-94 (2008).
73. Huang, Y. *et al.* Association between job strain and risk of incident stroke: A meta-analysis. *Neurology* **85**, 1648-54 (2015).
74. O'Donnell, M.J. *et al.* Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* **388**, 761-75 (2016).
75. Repetti, R.L., Taylor, S.E. & Seeman, T.E. Risky families: family social environments and the mental and physical health of offspring. *Psychol Bull* **128**, 330-66 (2002).

D.5 Supplementary Tables

Supplementary Table 1a The brief life events questionnaire based on The List of Threatening Experiences (LTE) self-reported by Generation Scotland participants. Participants asked to report stressful life events over last 6 months.

	Independent Stressful life Events
1	Did you suffer from a serious illness, injury or assault?
2	Did a serious illness, injury or assault happen to a close relative?
3	Did a parent, spouse (or partner), child, brother or sister of yours die?
4	Did a close family friend or relative die, such as an aunt, cousin or grandparent?
5	Was something you valued lost or stolen?
	Dependent Stressful life Events
6	Did you have a separation due to marital difficulties or break off a steady relationship?
7	Did you have a serious problem with a close friend, neighbour or relatives?
8	Were you made redundant or sacked from your job?
9	Were you seeking work without success for more than one month?
10	Did you have a major financial crisis such as losing the equivalent of three months income?
11	Did you have problems with the police involving a court appearance?
12	Did you/your wife or partner give birth to a child?*

* Item not classified into any category

Supplementary Table 1b Touchscreen questionnaire self-reported by UK Biobank participants to screen for SLE. ("illness, injury, bereavement, stress in last 2 years"; UK Biobank data-field 6145).

In the last 2 years have you experienced any of the following (you can select more than one answer):	
1	Serious illness, injury or assault to yourself.
2	Serious illness, injury or asssault of a close relative.
3	Death of a close relative.
4	Death of a spouse or partner.
5	Marital separation/divorce.
6	Financial dificulties.

Supplementary Table 2 The Patient Health Questionnaire used in UK Biobank. Participants asked to report over last two weeks.

Over the past two weeks:
1 Serious illness, injury or assault to yourself.
2 How often have you had little interest or pleasure in doing things?
3 How often have you felt tense, fidgety or restless?
4 How often have you felt tired or had little energy?

The General Health Questionnaire (GHQ) used in Generation Scotland consists of 28 items designed to identify whether an individual's current mental state differs from his/her typical state. This is used to identify the presence of symptoms compared to what is normal for the individua (over the past two weeks).

The General Health Questionnaire is copyrighted. The use of the questionnaire is licensed by GL Assessment (https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/). A license agreement must be completed beforehand and a user fee is required to all users (commercial and academic users).

Supplementary Table 3a. Genetic correlation with depressive symptoms estimated by bivariate REML using genome-wide complex trait analyisis (GCTA).

Trait	rG	SE	P	N
UK Biobank				
PHQ-SLE	0.723601	0.117723	**< 0.00001**	50000*
Generation Scotland				
GHQ-TSLE	1	0.83132	0.1145353	4919
GHQ-DSLE	1	0.630666	0.05644431	4919
GHQ-ISLE	1	69.248444	0.4942395	4919

*calculated in a random set of 50,000 individuals.
In UK Biobank, models were adjusted by age, sex and 15 PCs.
In Generation Scotland, models were adjusted by age, sex and 20 PCs.

Supplementary Table 3b. SNP heritability estimates from genome-wide complex trait analysis (GCTA)

Trait	SNP-heritability estimate	SE	P	N
UK Biobank				
PHQ	0.089858	0.005001	**< 0.00001**	99057
*PHQ (SLE adjusted)	0.078858	0.004928	**< 0.00001**	99057
SLE	0.040449	0.004667	**< 0.00001**	99057
Generation Scotland				
GHQ	0.071227	0.073081	0.1654	4919
^GHQ	0.13536	0.053511	**0.005146**	6751
TSLE	0.061389	0.071296	0.1895	4919
DSLE	0.130571	0.071161	**0.02879**	4919
ISLE	0.000001	0.072365	0.5	4919

In UK Biobank, models were adjusted by age, sex and 15PCs.
In GS, models were adjusted by age, sex and 20 PCs.
*SLE was added as a covariate.
^Heritability for GHQ was estimated in a sample of 6,751 unrelated individuals from the full GS cohort at baseline[1]. QC filters were applyied as described in the "Cohort description" at Materials & Methods section for the main sample of 4,919 but applying QC steps in the 21,525 GS individuals sample (instead of the 8,541 individials that reported updated measures as part of STRADL follow-up), and randomly selecting individuals when filtering by degree of relatedness (instead of maximizing retention of those individuals reporting higher number of SLE). SLE measures are not available for GS at baseline.

[1]Smith, B.H. et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol 42, 689-700 (2013).

Supplementary Table 4. Top GWAS results ($p<1x10^{-5}$).

SNP	CHR	BP	G/I	MAF	A1	A2	BETA	STAT	P	Location	Closest gene (±25kb)
GWAS of GHQ in Generation Scotland											
rs2381499	4	40211997	G	0.3454	A	G	0.05315	4.729	$2.32x10^{-6}$	Intronic	RHOH
rs16870827	5	70976503	G	0.1787	G	A	0.06384	4.492	$7.20\ x10^{-6}$	-	MCCC2(+21.97kb)
rs4552473	4	40205742	G	0.2803	A	C	0.05236	4.445	$8.98\ x10^{-6}$	Intronic	RHOH
GWAS of PHQ in UK Biobank											
rs7954341	12	81419397	I	0.3191	G	A	0.022	5.229	$1.71x10^{-7}$	-	-
rs10862219	12	81430043	I	0.3198	T	C	0.02163	5.144	$2.70x10^{-7}$	-	-
rs11114720	12	81443764	G	0.3175	A	G	0.02157	5.114	$3.15x10^{-7}$	-	-
rs960379	2	208088537	I	0.3961	C	T	0.02052	5.108	$3.27x10^{-7}$	Intronic	LOC101927865
rs7776980	7	117496319	I	0.3148	T	C	0.02186	5.054	$4.35x10^{-7}$	Intronic	CTTNBP2
rs11114723	12	81449343	I	0.317	A	C	0.02114	5.025	$5.03x10^{-7}$	-	ACSS3(-22.46kb)
rs17007930	12	81445762	I	0.317	T	C	0.0211	5.017	$5.26x10^{-7}$	-	-
rs72707252	5	1128891	G	0.02572	T	C	0.06152	4.981	$6.33x10^{-7}$	-	SLC12A7(+16.72kb)
rs2176912	12	81394117	I	0.3594	G	A	0.02039	4.958	$7.13x10^{-7}$	-	-
rs7782815	7	117585779	I	0.3122	G	A	0.02097	4.956	$7.21x10^{-7}$	-	-
rs17488728	7	117544180	I	0.4558	T	C	0.01946	4.939	$7.85x10^{-7}$	-	-
rs4275684	12	81465645	G	0.318	G	T	0.02075	4.921	$8.64x10^{-7}$	-	ACSS3(-6.163kb)
rs10272923	7	117572304	I	0.456	C	T	0.01929	4.898	$9.69x10^{-7}$	-	-
rs62471815	7	117582736	G	0.4561	A	G	0.01921	4.878	$1.07x10^{-6}$	-	-
rs11663824	18	50745236	I	0.3992	A	C	0.01944	4.833	$1.35x10^{-6}$	Intronic	DCC
rs1024434	7	117580053	I	0.4567	A	C	0.01905	4.827	$1.39x10^{-6}$	-	-
rs11695409	2	208088987	G	0.365	G	A	0.01946	4.797	$1.61x10^{-6}$	Intronic	LOC101927865
rs76549394	1	75935191	G	0.05128	G	A	-0.04215	-4.738	$2.16x10^{-6}$	Intronic	SLC44A5
rs4717577	7	70493364	G	0.4296	A	G	-0.01855	-4.7	$2.60x10^{-6}$	-	-
rs912337	13	99141272	I	0.4263	C	T	0.01862	4.683	$2.84x10^{-6}$	Intronic	STK24
rs10950237	7	70491251	G	0.3755	G	A	-0.01877	-4.648	$3.36x10^{-6}$	-	-
rs219174	2	33228292	I	0.4087	G	A	0.01843	4.615	$3.93x10^{-6}$	Intronic	LTBP1
rs60939828	18	50729267	G	0.4323	C	T	0.01826	4.602	$4.19x10^{-6}$	Intronic	DCC
rs6789329	3	136476862	I	0.359	C	T	0.01877	4.599	$4.25x10^{-6}$	-	STAG1(+5.617kb)

rs219158	2	33218037	I	0.3867	G	A	0.01849	4.568	$4.94x10^{-6}$	Intronic	LTBP1
rs847357	14	72780817	I	0.4749	G	A	0.01798	4.567	$4.95x10^{-6}$	Intronic	RGS6
rs2360675	2	208059640	I	0.4439	C	A	0.01818	4.562	$5.08x10^{-6}$	Intronic	LOC101927865
rs11114668	12	81366066	I	0.2337	G	A	0.02117	4.55	$5.37x10^{-6}$	-	-
rs8084351	18	50726559	I	0.4901	G	A	-0.0178	-4.536	$5.74x10^{-6}$	Intronic	DCC
rs74546496	11	109813282	G	0.01132	G	A	0.08419	4.523	$6.11x10^{-6}$	-	-
rs6928738	6	30931059	G	0.07183	T	C	0.03418	4.511	$6.47x10^{-6}$	-	DPCR1(+9.061kb)|MUC21(-20.42kb)
rs17488727	18	50760933	I	0.4212	C	A	0.01799	4.507	$6.59x10^{-6}$	Intronic	DCC
rs62100766	18	50737358	G	0.4009	T	G	0.01808	4.506	$6.61x10^{-6}$	Intronic	DCC
rs4917447	10	107438736	I	0.3211	C	T	0.01893	4.495	$6.98x10^{-6}$	-	-
rs917284	14	72737111	I	0.4229	G	T	-0.01785	-4.494	$7.01x10^{-6}$	Intronic	RGS6
rs9873767	3	135930406	G	0.2451	T	C	-0.02043	-4.489	$7.16x10^{-6}$	-	MSL2(+15.72kb)
rs6589381	11	113378952	I	0.4345	A	C	-0.01773	-4.484	$7.35x10^{-6}$	-	-
rs1352309	11	21911419	G	0.238	T	C	0.02062	4.48	$7.47x10^{-6}$	-	-
rs7232543	18	50718757	I	0.4362	G	A	0.01774	4.479	$7.52x10^{-6}$	Intronic	DCC
rs663895	11	78915746	G	0.4917	C	T	0.01749	4.477	$7.57x10^{-6}$	Intronic	TENM4
rs4372758	18	50683306	G	0.4429	C	T	0.01766	4.475	$7.66x10^{-6}$	Intronic	DCC
rs17487256	18	50718314	I	0.4015	T	C	0.0179	4.473	$7.74x10^{-6}$	Intronic	DCC
rs1568404	14	72736969	I	0.4682	T	C	0.01762	4.471	$7.80x10^{-6}$	Intronic	RGS6
rs9467626	6	25873746	G	0.1114	A	C	-0.02785	-4.471	$7.81x10^{-6}$	Intronic	SLC17A3
rs6785876	3	44621285	G	0.3227	A	G	0.0187	4.467	$7.94x10^{-6}$	Intronic	ZKSCAN7|ZNF660(-5.17kb)
rs4863713	4	140907254	I	0.3577	T	C	0.01828	4.463	$8.07x10^{-6}$	Intronic	MAML3
rs4842376	12	81437253	G	0.2312	C	T	0.02069	4.461	$8.17x10^{-6}$	-	-
rs10891540	11	113239082	I	0.4651	G	A	0.01758	4.458	$8.28x10^{-6}$	-	ANKK1(-19.43kb)|TTC12(+1.968kb)
rs16888206	5	57774538	I	0.09685	A	G	-0.02972	-4.445	$8.78x10^{-6}$	-	GAPT(-12.79kb)|PLK2(+18.57kb)
rs12727575	1	34052305	G	0.3741	T	C	0.0179	4.438	$9.11x10^{-6}$	Intronic	CSMD2
rs586818	11	78929880	G	0.4984	G	A	0.01732	4.432	$9.35x10^{-6}$	Intronic	TENM4
rs1538686	1	196397819	I	0.4133	G	A	-0.01773	-4.431	$9.37x10^{-6}$	Intronic	KCNT2
rs4479021	11	113383394	I	0.4351	A	G	-0.01757	-4.425	$9.65x10^{-6}$	-	-
rs219178	2	33231051	G	0.4095	G	A	0.01797	4.421	$9.83x10^{-6}$	Intronic	LTBP1

Supplementary Table 5. Prioritized genes from GWAS and GWEIS.

ensg: ENSG ID
Symbol: Gene Symbol
chr: Chromosome
start: Starting position of the gene
end: Ending position of the gene
strand: Strand of the gene
status: Status of the gene from Ensembl
type: Gene biotype from Ensembl
entrezID: entrez ID
HUGO: HUGO (HGNC) gene symbol
pLI: pLI score from ExAC database. The probability of being loss-of-function intolerant. The higher the score is, the more intolerant to loss-of-function mutations the gene is.
ncRVIS: Non-coding residual variation intolerance score. The higher the score is, the more intolerant to non-coding variation the gene is.
posMapSNPs: The number of SNPs mapped to gene based on positional mapping (after functional filtering if parameters are given).
posMapMaxCADD: The maximum CADD score of mapped SNPs by positional mapping.
eqtlMapSNPs: The number of SNPs mapped to the gene based on eQTL mapping.
ciMap: If the gene is mapped by chromatin interaction "Yes", otherwise "No".
ciMapts: Tissue/cell types of mapped chromatin interactions.
minGwasP: The minimum P-value of mapped SNPs.
IndSigSNPs: rsID of the all independent significant SNPs of mapped SNPs.
GenomicLocus: Index of genomic loci where mapped SNPs are from. This could contain more than one interval in the case that eQTLs are mapped to genes from distinct genomic risk loci.
In red prioritized genes identified by chromatin interaction mapping using Hi-C data.

ensg	symbol	chr	start	end	strand	status	type	entrezID	HUGO	pLI	ncRVIS	posMapSNPs	posMapMaxCADD	eqtlMapSNPs	ciMap	ciMapts	minGwasP	IndSigSNPs
UK Biobank GWAS PHQ																		
ENSG00000121904	CSMD2	1	33979609	34631443	-1	KNOWN	protein_coding	114784	CSMD2	NA	-0.899610176	6	2.526	0	No	NA	$9.11x10^{-6}$	rs12727575
ENSG00000162624	LHX8	1	75594119	75627218	1	KNOWN	protein_coding	431707	LHX8	0.663323206	-0.679265544	1	5.74	0	No	NA	NA	rs76549394
ENSG00000137968	SLC44A5	1	75667816	76076801	-1	KNOWN	protein_coding	204962	SLC44A5	0.03422869	0.661940842	9	6.207	0	No	NA	$2.16x10^{-6}$	rs76549394
ENSG00000162687	KCNT2	1	196194909	196578355	-1	KNOWN	protein_coding	343450	KCNT2	0.666802214	0.148619072	76	19.32	0	No	NA	$9.37x10^{-6}$	rs1538686
ENSG00000049323	LTBP1	2	33172039	33624576	1	KNOWN	protein_coding	4052	LTBP1	0.526234698	1.161890182	22	10.01	0	No	NA	$3.93x10^{-6}$	rs219174
ENSG00000118263	KLF7	2	207938861	208031991	-1	KNOWN	protein_coding	8609	KLF7	0.929930709	-2.115394599	10	17.65	0	No	NA	$2.31x10^{-5}$	rs960379
ENSG00000118260	CREB1	2	208394461	208468155	1	KNOWN	protein_coding	1385	CREB1	0.969960677	-0.523505755	0	0	0	Yes	Hippocampus	$2.31x10^{-5}$	rs960379
ENSG00000179152	TCAIM	3	44379611	44450943	1	KNOWN	protein_coding	285343	TCAIM	$4.46x10^{-5}$	0.31315397	3	6.739	0	No	NA	NA	rs6785876
ENSG00000185219	ZNF445	3	44481262	44519162	-1	KNOWN	protein_coding	353274	ZNF445	0.996478491	-1.777377016	17	8.948	0	No	NA	0.0002468	rs6785876

ENSG00000178917	ZNF852	3	44540462	44552128	-1	KNOWN	protein_coding	285346	ZNF852	$6.54x10^{-9}$	NA	5	3.349	0	No	NA	0.0001415	rs6785876
ENSG00000196345	ZKSCAN7	3	44596685	44624975	1	KNOWN	protein_coding	55888	ZKSCAN7	$1.18x10^{-5}$	-0.800781432	52	12.42	0	No	NA	$7.94x10^{-6}$	rs6785876
ENSG00000144792	ZNF660	3	44619715	44641186	1	KNOWN	protein_coding	285349	ZNF660	0.054805105	-1.658430892	36	14.13	0	No	NA	$7.94x10^{-6}$	rs6785876
ENSG00000186448	ZNF197	3	44626380	44689963	1	KNOWN	protein_coding	10168	ZNF197	$5.95x10^{-14}$	0.135916243	46	16.95	0	No	NA	$7.94x10^{-6}$	rs6785876
ENSG00000169981	ZNF35	3	44690219	44702283	1	KNOWN	protein_coding	7584	ZNF35	0.079290627	-0.732337005	6	16.95	0	No	NA	$1.39x10^{-5}$	rs6785876
ENSG00000196653	ZNF502	3	44754135	44765323	1	KNOWN	protein_coding	91392	ZNF502	$1.41x10^{-5}$	0.25289851	4	7.863	0	No	NA	NA	rs6785876
ENSG00000163808	KIF15	3	44803209	44914868	1	KNOWN	protein_coding	56992	KIF15	$7.47x10^{-13}$	-0.317071365	3	3.24	0	No	NA	$4.02x10^{-5}$	rs6785876
ENSG00000169964	TMEM42	3	44903361	44907162	1	KNOWN	protein_coding	131616	TMEM42	0.306977503	-0.038932273	3	3.24	0	No	NA	$4.02x10^{-5}$	rs6785876
ENSG00000163810	TGM4	3	44916100	44956482	1	KNOWN	protein_coding	7047	TGM4	$1.14x10^{-11}$	0.329751591	3	3.24	0	No	NA	$4.02x10^{-5}$	rs6785876
ENSG00000073711	PPP2R3A	3	135684515	135866733	1	KNOWN	protein_coding	5523	PPP2R3A	0.920487875	-0.863679102	10	5.037	0	No	NA	$3.13x10^{-5}$	rs6789329
ENSG00000174579	MSL2	3	135867764	135916083	-1	KNOWN	protein_coding	55167	MSL2	0.889663742	-1.742045868	2	3.974	0	Yes	Hippocampus	$4.25x10^{-6}$	rs6789329
ENSG00000114054	PCCB	3	135969148	136056738	1	KNOWN	protein_coding	5096	PCCB	0.000735839	-0.385160254	20	15.38	0	No	NA	$2.02x10^{-5}$	rs9873767;rs6789329
ENSG00000118007	STAG1	3	136055077	136471220	-1	KNOWN	protein_coding	10274	STAG1	0.99999991	-0.647637604	94	10.6	0	Yes	Hippocampus	$4.25x10^{-6}$	rs6789329
ENSG00000196782	MAML3	4	140637907	141075338	-1	KNOWN	protein_coding	55534	MAML3	0.329895048	-0.192465137	61	16.03	0	No	NA	$8.07x10^{-6}$	rs4863713
ENSG00000077063	CTTNBP2	7	117350705	117514193	-1	KNOWN	protein_coding	83992	CTTNBP2	$2.60x10^{-6}$	-0.293652714	22	21	0	No	NA	$4.35x10^{-7}$	rs7776980;rs17488728
ENSG00000149256	TENM4	11	78363876	79151992	-1	KNOWN	protein_coding	26011	TENM4	0.999718776	-0.693242157	3	5.233	0	No	NA	$7.57x10^{-6}$	rs663895
ENSG00000149292	TTC12	11	113185251	113254266	1	KNOWN	protein_coding	54970	TTC12	$1.45x10^{-7}$	-0.271528473	37	19.44	45	No	NA	$8.28x10^{-6}$	rs10891540
ENSG00000170209	ANKK1	11	113258513	113271140	1	KNOWN	protein_coding	255239	ANKK1	$2.51x10^{-8}$	-0.017548171	18	8.304	0	No	NA	$4.29x10^{-5}$	rs10891540
ENSG00000149295	DRD2	11	113280318	113346413	-1	KNOWN	protein_coding	1813	DRD2	0.733263889	NA	41	17.76	0	No	NA	$1.51x10^{-5}$	rs10891540;rs6589381
ENSG00000111052	LIN7A	12	81186299	81331704	-1	KNOWN	protein_coding	8825	LIN7A	0.041148496	-1.3739659	2	3.063	0	No	NA	NA	rs11114668
ENSG00000111058	ACSS3	12	81331594	81650533	1	KNOWN	protein_coding	79611	ACSS3	$3.00x10^{-11}$	-0.814138549	235	19.41	0	No	NA	$1.71x10^{-7}$	rs11114668;rs7954341;rs4842376
ENSG00000152767	FARP1	13	98794816	99102027	1	KNOWN	protein_coding	10160	FARP1	0.045263074	-0.444415571	1	4.372	0	No	NA	NA	rs912337
ENSG00000102572	STK24	13	99102455	99230194	-1	KNOWN	protein_coding	8428	STK24	0.797998297	0.417068263	55	9.214	0	No	NA	$2.84x10^{-6}$	rs912337
ENSG00000182732	RGS6	14	72399156	73030654	1	KNOWN	protein_coding	9628	RGS6	0.801361183	NA	106	16.13	0	No	NA	$4.95x10^{-6}$	rs917284;rs847357
ENSG00000187323	DCC	18	49866542	51057784	1	KNOWN	protein_coding	1630	DCC	0.999998957	-0.239950445	160	18.83	0	No	NA	$1.35x10^{-6}$	rs11663824

UK Biobank GWEIS - GxE effect

ENSG00000073282	TP63	3	189349205	189615068	1	KNOWN	protein_coding	8626	TP63	0.983222202	-0.204818685	25	9.698	0	No	NA	3.76x10-6	rs7650285
ENSG00000119471	HSDL2	9	115142217	115234690	1	KNOWN	protein_coding	84263	HSDL2	$3.95x10^{-5}$	-0.039855639	48	15.32	0	No	NA	5.86x10-5	rs2593684
ENSG00000230185	C9orf147	9	115196096	115249484	-1	KNOWN	protein_coding	100133204	C9orf147	NA	NA	25	15.32	0	No	NA	5.86x10-5	rs2593684
ENSG00000165185	KIAA1958	9	115249127	115431677	1	KNOWN	protein_coding	158405	KIAA1958	0.04900214	0.577176307	36	17.91	0	No	NA	3.60x10-6	rs2593684
ENSG00000139915	MDGA2	14	47308826	48144157	-1	KNOWN	protein_coding	161357	MDGA2	0.993216871	0.397758699	53	14.14	0	No	NA	8.44x10-6	rs11844549

ENSG00000154864	PIEZO2	18	10666480	11148587	-1	KNOWN	protein_coding	63895	PIEZO2	0.603955313	NA	1	1.29	0	No	NA	$9.68x10^{-6}$	rs77317628

UK Biobank GWEIS - joint effect

ENSG00000118263	KLF7	2	207938861	208031991	-1	KNOWN	protein_coding	8609	KLF7	0.929930709	-2.115394599	10	17.65	0	No	NA	0.000112279	rs960379
ENSG00000118260	CREB1	2	208394461	208468155	1	KNOWN	protein_coding	1385	CREB1	0.969960677	-0.523505755	0	0	0	Yes	Hippocampus	0.000112279	rs960379
ENSG00000185129	PURA	5	139487362	139496321	1	KNOWN	protein_coding	5813	PURA	0.845099484	-0.96108256	8	13.86	0	No	NA	$5.99x10^{-6}$	rs2974421
ENSG00000182700	IGIP	5	139505521	139508391	1	KNOWN	protein_coding	492311	IGIP	0.419918844	-1.738238174	4	13.86	0	No	NA	$5.99x10^{-6}$	rs2974421
ENSG00000112232	KHDRBS2	6	62389865	62996132	-1	KNOWN	protein_coding	202559	KHDRBS2	0.142878302	-0.85475727	1	3.892	0	No	NA	NA	rs77225982
ENSG00000127928	GNGT1	7	93220885	93540577	1	KNOWN	protein_coding	2792	GNGT1	0.276892737	0.237481529	16	8.068	0	No	NA	$4.46x10^{-6}$	rs2724063
ENSG00000184903	IMMP2L	7	110303110	111202573	-1	KNOWN	protein_coding	83943	IMMP2L	0.040760918	-0.338556389	22	16.6	0	No	NA	$3.52x10^{-6}$	rs2091309
ENSG00000188938	FAM120AOS	9	96208776	96215874	-1	KNOWN	protein_coding	158293	FAM120AOS	0.514390707	0.007431442	7	4.179	0	No	NA	$7.03x10^{-5}$	rs10116422
ENSG00000048828	FAM120A	9	96214004	96328397	1	KNOWN	protein_coding	23196	FAM120A	0.99990705	-0.298117336	86	14.9	0	No	NA	$3.25x10^{-5}$	rs10116422
ENSG00000197724	PHF2	9	96338689	96441869	1	KNOWN	protein_coding	5253	PHF2	0.994132137	0.745197979	43	7.829	0	No	NA	$8.08x10^{-6}$	rs10116422
ENSG00000197467	COL13A1	10	71561644	71724031	1	KNOWN	protein_coding	1305	COL13A1	0.003928968	0.409844169	3	4.856	0	No	NA	$7.70x10^{-6}$	rs7070915
ENSG00000111058	ACSS3	12	81331594	81650533	1	KNOWN	protein_coding	79611	ACSS3	$3.00x10^{-11}$	-0.814138549	137	13.83	0	No	NA	$4.37x10^{-7}$	rs7954341
ENSG00000166748	AGBL1	15	86685227	87572283	1	KNOWN	protein_coding	123624	AGBL1	$3.60x10^{-23}$	-0.103443516	269	22.1	0	No	NA	$2.35x10^{-6}$	rs8038215
ENSG00000090863	GLG1	16	74485856	74641012	-1	KNOWN	protein_coding	2734	GLG1	0.997711424	-0.205896764	10	21.1	0	No	NA	$5.15x10^{-6}$	rs77995020

Generation Scotland GWEIS using TSLE as exposure - GxE effect

ENSG00000154305	MIA3	1	222791428	222841354	1	KNOWN	protein_coding	375056	MIA3	0.471388106	-0.616197151	22	14.33	0	No	NA	NA	rs17163441
ENSG00000186063	AIDA	1	222841355	222886552	-1	KNOWN	protein_coding	64853	AIDA	0.082372382	-1.192388865	27	14.95	0	No	NA	$4.98x10^{-6}$	rs17163441
ENSG00000162819	BROX	1	222885895	222908538	1	KNOWN	protein_coding	148362	BROX	$2.48x10^{-5}$	1.32122969	17	9.915	0	No	NA	$4.98x10^{-6}$	rs17163441
ENSG00000197520	FAM177B	1	222910549	222924147	1	KNOWN	protein_coding	400823	FAM177B	$9.21x10^{-6}$	-0.425513654	5	10.91	0	No	NA	$5.89x10^{-6}$	rs17163441
ENSG00000137275	RIPK1	6	3064225	3115421	1	KNOWN	protein_coding	8737	RIPK1	0.4181774	-0.203587189	3	3.669	0	No	NA	$5.06x10^{-6}$	rs17548315
ENSG00000137274	BPHL	6	3118608	3153812	1	KNOWN	protein_coding	670	BPHL	0.000194971	1.195504957	2	4.684	0	No	NA	$5.06x10^{-6}$	rs17548315
ENSG00000137267	TUBB2A	6	3153903	3157760	-1	KNOWN	protein_coding	7280	TUBB2A	NA	NA	1	4.684	0	No	NA	NA	rs17548315
ENSG00000111816	FRK	6	116252312	116381921	-1	KNOWN	protein_coding	2444	FRK	$6.59x10^{-10}$	2.58876413	2	4.816	0	No	NA	NA	rs11754507
ENSG00000178425	NT5DC1	6	116422012	116570660	1	KNOWN	protein_coding	221294	NT5DC1	0.326411116	-0.101658908	1	7.132	0	Yes	Hippocampus	NA	rs11754507
ENSG00000123500	COL10A1	6	116440086	116479910	-1	KNOWN	protein_coding	1300	COL10A1	$6.36x10^{-5}$	-0.141903212	1	7.132	0	No	NA	NA	rs11754507
ENSG00000111817	DSE	6	116575336	116762424	1	KNOWN	protein_coding	29940	DSE	0.385415297	1.399470369	1	2.355	0	No	NA	$1.04x10^{-7}$	rs11754507
ENSG00000173626	TRAPPC3L	6	116816152	116866773	-1	KNOWN	protein_coding	100128327	TRAPPC3L	NA	-0.063711938	3	2.09	0	No	NA	NA	rs11754507
ENSG00000178033	FAM26E	6	116832809	116845955	1	KNOWN	protein_coding	254228	FAM26E	0.001108388	2.210740591	1	1.557	0	No	NA	NA	rs11754507
ENSG00000164451	FAM26D	6	116850174	116880031	1	KNOWN	protein_coding	221301	FAM26D	0.033838822	0.383426399	4	2.09	0	No	NA	NA	rs11754507
ENSG00000184672	RALYL	8	85095022	85834079	1	KNOWN	protein_coding	138046	RALYL	0.029690036	0.669570867	6	18.19	0	No	NA	$1.07x10^{-6}$	rs12677170

Generation Scotland GWEIS using TSLE as exposure - joint effect

ENSG00000160145	KALRN	3	123798870	124445172	1	KNOWN	protein_coding	8997	KALRN	0.999996915	-0.64291779	1	5.012	0	No	NA	$5.69x10^{-6}$	rs2289841
ENSG00000137275	RIPK1	6	3064225	3115421	1	KNOWN	protein_coding	8737	RIPK1	0.4181774	-0.203587189	3	3.669	0	No	NA	$8.18x10^{-6}$	rs17548315
ENSG00000137274	BPHL	6	3118608	3153812	1	KNOWN	protein_coding	670	BPHL	0.000194971	1.195504957	2	4.684	0	No	NA	$8.18x10^{-6}$	rs17548315
ENSG00000137267	TUBB2A	6	3153903	3157760	-1	KNOWN	protein_coding	7280	TUBB2A	NA	NA	1	4.684	0	No	NA	NA	rs17548315
ENSG00000111816	FRK	6	116252312	116381921	-1	KNOWN	protein_coding	2444	FRK	$6.59x10^{-10}$	2.58876413	2	4.816	0	No	NA	NA	rs11754507
ENSG00000178425	NT5DC1	6	116422012	116570660	1	KNOWN	protein_coding	221294	NT5DC1	0.326411116	-0.101658908	1	7.132	0	Yes	Hippocampus	NA	rs11754507
ENSG00000123500	COL10A1	6	116440086	116479910	-1	KNOWN	protein_coding	1300	COL10A1	$6.36x10^{-5}$	-0.141903212	1	7.132	0	No	NA	NA	rs11754507
ENSG00000111817	DSE	6	116575336	116762424	1	KNOWN	protein_coding	29940	DSE	0.385415297	1.399470369	1	2.355	0	No	NA	$3.50x10^{-7}$	rs11754507
ENSG00000173626	TRAPPC3L	6	116816152	116866773	-1	KNOWN	protein_coding	100128327	TRAPPC3L	NA	-0.063711938	3	2.09	0	No	NA	NA	rs11754507
ENSG00000178033	FAM26E	6	116832809	116845955	1	KNOWN	protein_coding	254228	FAM26E	0.001108388	2.210740591	1	1.557	0	No	NA	NA	rs11754507
ENSG00000164451	FAM26D	6	116850174	116880031	1	KNOWN	protein_coding	221301	FAM26D	0.033838822	0.383426399	4	2.09	0	No	NA	NA	rs11754507
ENSG00000170786	SDR16C5	8	57212569	57233335	-1	KNOWN	protein_coding	195814	SDR16C5	0.00291744	0.076651261	1	1.231	0	No	NA	$5.31x10^{-6}$	rs4409393
ENSG00000184672	RALYL	8	85095022	85834079	1	KNOWN	protein_coding	138046	RALYL	0.029690036	0.669570867	6	18.19	0	No	NA	$3.66x10^{-6}$	rs12677170
ENSG00000151474	FRMD4A	10	13685706	14504141	-1	KNOWN	protein_coding	55691	FRMD4A	0.999986541	2.100618864	13	5.918	0	No	NA	$6.01x10^{-6}$	rs11259022

Generation Scotland GWEIS using DSLE as exposure - GxE effect

ENSG00000111816	FRK	6	116252312	116381921	-1	KNOWN	protein_coding	2444	FRK	$6.59x10^{-10}$	2.58876413	2	4.816	0	No	NA	NA	rs11754507
ENSG00000178425	NT5DC1	6	116422012	116570660	1	KNOWN	protein_coding	221294	NT5DC1	0.326411116	-0.101658908	1	7.132	0	Yes	Hippocampus	NA	rs11754507
ENSG00000123500	COL10A1	6	116440086	116479910	-1	KNOWN	protein_coding	1300	COL10A1	$6.36x10^{-5}$	-0.141903212	1	7.132	0	No	NA	NA	rs11754507
ENSG00000111817	DSE	6	116575336	116762424	1	KNOWN	protein_coding	29940	DSE	0.385415297	1.399470369	1	2.355	0	No	NA	$2.11x10^{-6}$	rs11754507
ENSG00000173626	TRAPPC3L	6	116816152	116866773	-1	KNOWN	protein_coding	100128327	TRAPPC3L	NA	-0.063711938	3	2.09	0	No	NA	NA	rs11754507
ENSG00000178033	FAM26E	6	116832809	116845955	1	KNOWN	protein_coding	254228	FAM26E	0.001108388	2.210740591	1	1.557	0	No	NA	NA	rs11754507
ENSG00000164451	FAM26D	6	116850174	116880031	1	KNOWN	protein_coding	221301	FAM26D	0.033838822	0.383426399	4	2.09	0	No	NA	NA	rs11754507
ENSG00000148468	FAM171A1	10	15253642	15413061	-1	KNOWN	protein_coding	221061	FAM171A1	0.987016298	-1.169213337	24	8.699	0	No	NA	$5.35x10^{-6}$	rs11259593
ENSG00000134640	MTNR1B	11	92702886	92718232	1	KNOWN	protein_coding	4544	MTNR1B	0.00870273	-0.165792302	31	11.66	0	No	NA	$1.76x10^{-6}$	rs6483212

Generation Scotland GWEIS using DSLE as exposure - joint effect

ENSG00000111816	FRK	6	116252312	116381921	-1	KNOWN	protein_coding	2444	FRK	$6.59x10^{-10}$	2.58876413	2	4.816	0	No	NA	NA	rs11754507
ENSG00000178425	NT5DC1	6	116422012	116570660	1	KNOWN	protein_coding	221294	NT5DC1	0.326411116	-0.101658908	1	7.132	0	Yes	Hippocampus	NA	rs11754507
ENSG00000123500	COL10A1	6	116440086	116479910	-1	KNOWN	protein_coding	1300	COL10A1	$6.36x10^{-5}$	-0.141903212	1	7.132	0	No	NA	NA	rs11754507
ENSG00000111817	DSE	6	116575336	116762424	1	KNOWN	protein_coding	29940	DSE	0.385415297	1.399470369	1	2.355	0	No	NA	$3.34x10^{-6}$	rs11754507
ENSG00000173626	TRAPPC3L	6	116816152	116866773	-1	KNOWN	protein_coding	100128327	TRAPPC3L	NA	-0.063711938	3	2.09	0	No	NA	NA	rs11754507
ENSG00000178033	FAM26E	6	116832809	116845955	1	KNOWN	protein_coding	254228	FAM26E	0.001108388	2.210740591	1	1.557	0	No	NA	NA	rs11754507

ENSG00000164451	FAM26D	6	116850174	116880031	1	KNOWN	protein_coding	221301	FAM26D	0.033838822	0.383426399	4	2.09	0	No	NA	NA	rs11754507
ENSG00000170786	SDR16C5	8	57212569	57233335	-1	KNOWN	protein_coding	195814	SDR16C5	0.00291744	0.076651261	1	1.231	0	No	NA	$1.62x10^{-6}$	rs4409393
ENSG00000099810	MTAP	9	21802542	21931646	1	KNOWN	protein_coding	4507	MTAP	0.000179316	1.589224827	30	17.08	0	No	NA	$8.56x10^{-6}$	rs10811635
ENSG00000151474	FRMD4A	10	13685706	14504141	-1	KNOWN	protein_coding	55691	FRMD4A	0.999986541	2.100618864	13	5.918	0	No	NA	$4.64x10^{-6}$	rs11259022
ENSG00000134640	MTNR1B	11	92702886	92718232	1	KNOWN	protein_coding	4544	MTNR1B	0.00870273	-0.165792302	31	11.66	0	No	NA	$5.52x10^{-6}$	rs6483212
Generation Scotland GWEIS using ISLE as exposure - GxE effect																		
ENSG00000137275	RIPK1	6	3064225	3115421	1	KNOWN	protein_coding	8737	RIPK1	0.4181774	-0.203587189	3	3.669	0	No	NA	$2.57x10^{-6}$	rs17548315
ENSG00000137274	BPHL	6	3118608	3153812	1	KNOWN	protein_coding	670	BPHL	0.000194971	1.195504957	2	4.684	0	No	NA	$2.57x10^{-6}$	rs17548315
ENSG00000137267	TUBB2A	6	3153903	3157760	-1	KNOWN	protein_coding	7280	TUBB2A	NA	NA	1	4.684	0	No	NA	NA	rs17548315
ENSG00000186472	PCLO	7	82383329	82792246	-1	KNOWN	protein_coding	27445	PCLO	1	-0.00757724	1	5.158	0	No	NA	NA	rs10250565
ENSG00000184672	RALYL	8	85095022	85834079	1	KNOWN	protein_coding	138046	RALYL	0.029690036	0.669570867	6	18.19	0	No	NA	$2.09x10^{-6}$	rs12677170
ENSG00000053918	KCNQ1	11	2465914	2870339	1	KNOWN	protein_coding	3784	KCNQ1	$2.45x10^{-5}$	1.153237047	3	5.075	0	No	NA	$4.40x10^{-6}$	rs800336
Generation Scotland GWEIS using ISLE as exposure - joint effect																		
ENSG00000064692	SNCAIP	5	121647049	121799914	1	KNOWN	protein_coding	9627	SNCAIP	0.000362968	NA	40	15.46	0	Yes	Hippocampus	$5.08x10^{-6}$	rs2242223
ENSG00000137275	RIPK1	6	3064225	3115421	1	KNOWN	protein_coding	8737	RIPK1	0.4181774	-0.203587189	3	3.669	0	No	NA	$3.89x10^{-6}$	rs17548315
ENSG00000137274	BPHL	6	3118608	3153812	1	KNOWN	protein_coding	670	BPHL	0.000194971	1.195504957	2	4.684	0	No	NA	$3.89x10^{-6}$	rs17548315
ENSG00000137267	TUBB2A	6	3153903	3157760	-1	KNOWN	protein_coding	7280	TUBB2A	NA	NA	1	4.684	0	No	NA	NA	rs17548315
ENSG00000184672	RALYL	8	85095022	85834079	1	KNOWN	protein_coding	138046	RALYL	0.029690036	0.669570867	6	18.19	0	No	NA	$5.27x10^{-6}$	rs12677170

Supplementary Table 6. Gene set enriched by genes prioritized by FUMA from UK Biobank GWAS PHQ

N: The total number of genes in the gene set.
n: The total number of prioritized genes in the gene set.

UK Biobank GWAS of PHQ

GeneSet	N	n	*p*-value	adjusted *p*	genes
All Canonical Pathways					
kegg propanoate metabolism	32	2	2.40×10^{-5}	2.59×10^{-2}	ACSS3, PCCB
reactome generic transcription pathway	340	4	3.02×10^{-4}	2.59×10^{-2}	ZNF445, ZKSCAN7, ZNF197, MAML3
GO biological processes					
positive regulation of long term synaptic potentiation	14	2	1.80×10^{-6}	8.37×10^{-3}	DRD2, CREB1
regulation of long term synaptic potentiation	19	2	4.76×10^{-6}	1.05×10^{-2}	DRD2, CREB1
central nervous system neuron development	70	3	6.80×10^{-6}	1.05×10^{-2}	LHX8, DRD2, DCC
neuron differentiation	870	8	1.41×10^{-5}	1.64×10^{-2}	LHX8, TENM4, DRD2, FARP1, DCC, KLF7, CREB1, PPP2R3A
neuron development	683	7	1.91×10^{-5}	1.77×10^{-2}	LHX8, TENM4, DRD2, FARP1, DCC, KLF7, CREB1
positive regulation of multicellular organism growth	32	2	2.40×10^{-5}	1.86×10^{-2}	DRD2, CREB1
forebrain neuron development	34	2	2.89×10^{-5}	1.92×10^{-2}	LHX8, DRD2
cell morphogenesis involved in neuron differentiation	367	5	3.91×10^{-5}	2.25×10^{-2}	DRD2, FARP1, DCC, KLF7, CREB1
dendrite morphogenesis	42	2	5.49×10^{-5}	2.25×10^{-2}	FARP1, KLF7
pituitary gland development	42	2	5.49×10^{-5}	2.25×10^{-2}	DRD2, CREB1
neuron projection morphogenesis	400	5	6.31×10^{-5}	2.25×10^{-2}	DRD2, FARP1, DCC, KLF7, CREB1
regulation of phosphatase activity	125	3	6.71×10^{-5}	2.25×10^{-2}	DRD2, FARP1, PPP2R3A
visual behavior	50	2	9.28×10^{-5}	2.25×10^{-2}	DRD2, CREB1
response to nicotine	51	2	9.84×10^{-5}	2.25×10^{-2}	DRD2, CREB1
neurogenesis	1396	9	1.00×10^{-4}	2.25×10^{-2}	LHX8, TENM4, DRD2, FARP1, STK24, DCC, KLF7, CREB1, PPP2R3A
regulation of developmental growth	286	4	1.36×10^{-4}	2.25×10^{-2}	DRD2, DCC, CREB1, PPP2R3A
regulation of phosphoprotein phosphatase activity	57	2	1.37×10^{-4}	2.25×10^{-2}	DRD2, PPP2R3A
regulation of dephosphorylation	155	3	1.54×10^{-4}	2.25×10^{-2}	DRD2, FARP1, PPP2R3A

central nervous system neuron differentiation	166	3	$2.01x10^{-4}$	$2.25x10^{-2}$	LHX8, DRD2, DCC
forebrain generation of neurons	66	2	$2.12x10^{-4}$	$2.25x10^{-2}$	LHX8, DRD2
regulation of multicellular organism growth	66	2	$2.12x10^{-4}$	$2.25x10^{-2}$	DRD2, CREB1
synapse assembly	67	2	$2.22x10^{-4}$	$2.25x10^{-2}$	DRD2, FARP1
cell morphogenesis involved in differentiation	512	5	$2.43x10^{-4}$	$2.25x10^{-2}$	DRD2, FARP1, DCC, KLF7, CREB1
associative learning	73	2	$2.86x10^{-4}$	$2.34x10^{-2}$	DRD2, CREB1
diencephalon development	76	2	$3.22x10^{-4}$	$2.40x10^{-2}$	DRD2, CREB1
neuron projection development	542	5	$3.30x10^{-4}$	$2.40x10^{-2}$	DRD2, FARP1, DCC, KLF7, CREB1
dendrite development	79	2	$3.61x10^{-4}$	$2.40x10^{-2}$	FARP1, KLF7
neuron projection guidance	204	3	$4.40x10^{-4}$	$2.61x10^{-2}$	DCC, KLF7, CREB1
regulation of gliogenesis	90	2	$5.30x10^{-4}$	$2.93x10^{-2}$	TENM4, CREB1
central nervous system development	865	6	$6.63x10^{-4}$	$3.24x10^{-2}$	LHX8, TENM4, DRD2, DCC, CREB1, CTTNBP2
memory	98	2	$6.79x10^{-4}$	$3.24x10^{-2}$	DRD2, CREB1
cell part morphogenesis	631	5	$7.38x10^{-4}$	$3.40x10^{-2}$	DRD2, FARP1, DCC, KLF7, CREB1
regulation of circadian rhythm	102	2	$7.63x10^{-4}$	$3.44x10^{-2}$	DRD2, CREB1
cellular component morphogenesis	898	6	$8.28x10^{-4}$	$3.47x10^{-2}$	TENM4, DRD2, FARP1, DCC, KLF7, CREB1
cognition	250	3	$9.42x10^{-4}$	$3.64x10^{-2}$	LHX8, DRD2, CREB1
neuron migration	110	2	$9.50x10^{-4}$	$3.64x10^{-2}$	DRD2, DCC
regulation of embryonic development	114	2	$1.05x10^{-3}$	$3.80x10^{-2}$	TENM4, PPP2R3A
positive regulation of hormone secretion	117	2	$1.14x10^{-3}$	$4.00x10^{-2}$	DRD2, CREB1
endocrine system development	123	2	$1.31x10^{-3}$	$4.30x10^{-2}$	DRD2, CREB1
regulation of g protein coupled receptor protein signaling pathway	126	2	$1.41x10^{-3}$	$4.42x10^{-2}$	DRD2, RGS6
learning	130	2	$1.54x10^{-3}$	$4.62x10^{-2}$	DRD2, CREB1
GO cellular components					
neuron spine	121	3	$5.91x10^{-5}$	$3.45x10^{-2}$	DRD2, FARP1, CTTNBP2
GO molecular functions					
nucleic acid binding transcription factor activity	1180	8	$1.52x10^{-4}$	$3.75x10^{-2}$	KLF7, CREB1, ZNF445, ZKSCAN7, ZNF660, ZNF197, ZNF35, STAG1
GWAS catalog					

Age-related cataracts (age at onset)	3	2	5.01×10^{-9}	5.39×10^{-6}	LIN7A, ACSS3
Brain structure	16	2	2.76×10^{-6}	1.49×10^{-3}	CSMD2, FARP1
Schizophrenia	450	6	1.12×10^{-5}	3.89×10^{-3}	DRD2, RGS6, PPP2R3A, MSL2, PCCB, STAG1
Response to amphetamines	33	2	2.64×10^{-5}	4.73×10^{-3}	CSMD2, FARP1
Fibrinogen	41	2	5.10×10^{-5}	6.10×10^{-3}	MSL2, PCCB
Acne (severe)	160	3	1.74×10^{-4}	7.82×10^{-3}	KIF15, TMEM42, TGM4
Fibrinogen levels	67	2	2.22×10^{-4}	8.64×10^{-3}	PPP2R3A, PCCB
Educational attainment	72	2	2.75×10^{-4}	9.25×10^{-3}	FARP1, STK24
Waist circumference adjusted for body mass index	87	2	4.80×10^{-4}	1.36×10^{-2}	LTBP1, PPP2R3A

Supplementary Table 7. Top GWEIS results in UKB ($p<1x10^{-5}$)

SNP	CHR	BP	G/I	MAF	A1	A2	BETA	*p*-value	Location	Closest gene (<25kb)
GxE effect										
rs582813	18	25079009	G	0.4665	C	T	0.0283718	$1.13x10^{-7}$	-	-
rs60757853	8	109854124	G	0.2061	T	C	0.0329474	$3.79x10^{-7}$	-	-
rs2671750	3	5059576	G	0.1597	T	C	0.0359233	$5.66x10^{-7}$	-	-
rs13253194	8	109857400	I	0.2058	A	G	0.0324246	$6.06x10^{-7}$	-	-
rs2593684	9	115347088	I	0.251	A	G	-0.02881	$3.60x10^{-6}$	Intronic	KIAA1958
rs7650285	3	189500067	I	0.1107	A	G	0.0388284	$3.76x10^{-6}$	Intronic	TP63
rs2480061	1	14808756	I	0.4092	G	T	0.0244273	$6.64x10^{-6}$	-	-
rs12630584	3	189498809	I	0.1106	A	G	0.0375951	$7.60x10^{-6}$	Intronic	TP63
rs16864784	3	189499173	G	0.1108	G	A	0.0375224	$8.29x10^{-6}$	Intronic	TP63
rs9831427	3	142955098	I	0.3298	C	A	0.0253031	$8.31x10^{-6}$	-	-
rs11844549	14	47310983	I	0.4905	T	C	0.0239575	$8.44x10^{-6}$	Exonic	MDGA2
rs77317628	18	10900786	G	0.06402	T	C	-0.0481837	$9.68x10^{-6}$	Intronic	PIEZO2
rs4979128	9	115289605	I	0.2497	C	T	-0.0274627	$9.97x10^{-6}$	Intronic	KIAA1958
Joint effect										
rs7954341	12	81419397	I	0.3191	G	A	0.02107527	$4.37x10^{-7}$	-	-
rs582813	18	25079009	G	0.4665	C	T	0.0117316	$5.84x10^{-7}$	-	-
rs10862219	12	81430043	I	0.3198	T	C	0.02079495	$5.95x10^{-7}$	-	-
rs11114720	12	81443764	G	0.3175	A	G	0.02048935	$7.19x10^{-7}$	-	-
rs11114723	12	81449343	I	0.317	A	C	0.01993236	$1.00x10^{-6}$	-	ACSS3(-22.46kb)
rs17007930	12	81445762	I	0.317	T	C	0.01989167	$1.04x10^{-6}$	-	-
rs2176912	12	81394117	I	0.3594	G	A	0.01857864	$1.35x10^{-6}$	-	-
rs960379	2	208088537	I	0.3961	C	T	0.02318179	$1.39x10^{-6}$	Intronic	LOC101927865
rs4275684	12	81465645	G	0.318	G	T	0.01933315	$1.80x10^{-6}$	-	ACSS3(-6.163kb)
rs60757853	8	109854124	G	0.2061	T	C	0.0135463	$1.94x10^{-6}$	-	-
rs2671750	3	5059576	G	0.1597	T	C	0.0133213	$2.15x10^{-6}$	-	-
rs8038215	15	87538753	G	0.05269	C	T	0.0027143	$2.35x10^{-6}$	Intronic	AGBL1
rs13253194	8	109857400	I	0.2058	A	G	0.0136343	$3.20x10^{-6}$	-	-
rs2091309	7	110376327	I	0.3564	G	A	-0.02377593	$3.52x10^{-6}$	Intronic	IMMP2L

rs7239568	18	51964773	G	0.3341	C	A	-0.01611355	$4.03x10^{-6}$	-	-
rs2974421	5	139470240	I	0.1777	G	A	0.0090438	$4.11x10^{-6}$	-	LINC01024(-12.27kb)|PURA(-23.47kb)
rs2724063	7	93401210	I	0.4467	A	G	-0.022144542	$4.46x10^{-6}$	-	-
rs77995020	16	74510544	G	0.0924	A	C	-0.0392394	$5.15x10^{-6}$	Intronic	GLG1
rs10060173	5	139501147	I	0.1807	T	C	0.0085887	$5.99x10^{-6}$	-	IGIP(-4.373kb)|LINC01024(+13.55kb)|PURA(+2.146kb)
rs77225982	6	62328533	G	0.04729	G	A	-0.0252879	$6.00x10^{-6}$	-	-
rs2429310	5	139508705	G	0.1817	A	G	0.0080097	$6.15x10^{-6}$	-	IGIP(+0.314kb)|LINC01024(+21.11kb)|PURA(+9.704kb)
rs7070915	10	71652128	I	0.4949	C	T	0.02339642	$7.70x10^{-6}$	Intronic	COL13A1
rs114608878	6	62266630	G	0.04717	T	C	-0.0247592	$8.00x10^{-6}$	-	MTRNR2L9(-17.38kb)
rs11852282	15	87555349	G	0.04482	A	G	-0.0057676	$8.02x10^{-6}$	Intronic	AGBL1
rs10116422	9	96357739	I	0.3869	C	A	-0.02136075	$8.08x10^{-6}$	Intronic	PHF2
rs4917447	10	107438736	I	0.3211	C	T	0.0240048	$8.28x10^{-6}$	-	-
rs10950237	7	70491251	G	0.3755	G	A	-0.01656725	$8.44x10^{-6}$	-	-
rs7908373	10	107479864	I	0.3008	T	C	0.0246885	$8.68x10^{-6}$	-	-
rs2881642	15	87560258	I	0.04366	A	G	-0.003438	$9.62x10^{-6}$	Intronic	AGBL1
rs4376443	7	110391589	I	0.3384	T	C	-0.02346326	$9.80x10^{-6}$	Intronic	IMMP2L

Supplementary Table 8. Top TSLE GWEIS results in Generation Scotland ($p<1x10^{-5}$) - Total SLE as exposure

SNP	CHR	BP	MAF	A1	A2	BETA	*p*-value	Location	Closest gene (<25kb)
GxE effect									
rs12789145	**11**	**94407736**	**0.08002**	**C**	**A**	**0.059937**	**$4.95x10^{-9}$**	**-**	**-**
rs11754507	6	116697130	0.01701	G	A	0.126854	$1.04x10^{-7}$	Intronic	DSE
rs17070072	18	60205542	0.3437	A	C	-0.0361116	$9.02x10^{-7}$	Intronic	ZCCHC2
rs12677170	8	85446567	0.1169	C	A	-0.0539766	$1.07x10^{-6}$	Intronic	RALYL
rs10003191	4	17336925	0.2646	G	A	-0.0347034	$2.24x10^{-6}$	-	-
rs2059710	2	60838156	0.1609	A	G	0.0393715	$2.53x10^{-6}$	-	-
rs614123	13	34946870	0.291	A	G	0.0367452	$3.74x10^{-6}$	-	-
rs735133	2	60842571	0.1552	C	A	0.0392745	$4.67x10^{-6}$	-	-
rs17163441	1	222882892	0.1436	C	A	-0.0509662	$4.98x10^{-6}$	Intronic	AIDA
rs17548315	6	3114947	0.01824	A	G	0.0700267	$5.06x10^{-6}$	Exonic	RIPK1
rs17163437	1	222878687	0.1439	A	C	-0.0507477	$5.37x10^{-6}$	Intronic	AIDA
rs4364235	4	174635674	0.4482	A	G	-0.0319482	$5.46x10^{-6}$	-	-
rs3748631	1	222886588	0.144	A	G	-0.050744	$5.52x10^{-6}$	Intronic	BROX
rs12058171	1	222913107	0.1439	G	A	-0.0505684	$5.89x10^{-6}$	Exonic	FAM177B
rs6502321	17	13855742	0.07297	A	G	0.0520313	$6.69x10^{-6}$	-	-
rs9952353	18	19672632	0.07297	G	A	-0.0387656	$9.29x10^{-6}$	-	-
Joint effect									
rs12789145	11	94407736	0.08002	C	A	0.0005347	$2.77x10^{-8}$	-	-
rs11754507	6	116697130	0.01701	G	A	-0.021379	$3.50x10^{-7}$	Intronic	DSE
rs9952353	18	19672632	0.07297	G	A	0.0388734	$8.04x10^{-7}$	-	-
rs1579282	5	155074865	0.3308	A	G	-0.03026209	$1.03x10^{-6}$	-	-
rs17070072	18	60205542	0.3437	A	C	-0.0123343	$1.72x10^{-6}$	Intronic	ZCCHC2
rs10164188	18	11444650	0.02873	A	G	-0.08424123	$1.86x10^{-6}$	-	-
rs614123	13	34946870	0.291	A	G	-0.0266327	$2.43x10^{-6}$	-	-
rs12677170	8	85446567	0.1169	C	A	-0.0028139	$3.66x10^{-6}$	Intronic	RALYL
rs17244735	13	73104490	0.04892	G	A	0.0995494	$3.81x10^{-6}$	-	-
rs7577053	2	49849708	0.2334	A	G	-0.02605342	$4.96x10^{-6}$	-	-
rs11663716	18	19704036	0.2645	G	A	0.03751	$5.08x10^{-6}$	-	-

rs4409393	8	57214372	0.3665	A	G	-0.050243	$5.31x10^{-6}$	Intronic	SDR16C5
rs2289841	3	124209438	0.3568	A	G	0.0506544	$5.69x10^{-6}$	Intronic	KALRN
rs11259022	10	14385164	0.1712	A	G	0.06630978	$6.01x10^{-6}$	Intronic	FRMD4A(+12.3kb)
rs4800108	18	19694044	0.3445	C	A	0.033282	$7.39x10^{-6}$	-	-
rs1010811	18	19677786	0.2512	G	A	0.0352367	$7.82x10^{-6}$	-	-
rs17548315	6	3114947	0.01824	A	G	0.0305173	$8.18x10^{-6}$	Exonic	RIPK1
rs7736123	5	33039387	0.05975	A	G	0.0288125	$8.68x10^{-6}$	-	-

Supplementary Table 9. Top DSLE GWEIS results in Generation Scotland ($p<1x10^{-5}$) - Dependent SLE as exposure

SNP	CHR	BP	MAF	A1	A2	BETA	*p*-value	Location	Closest gene (<25kb)
GxE effect									
rs17070072	**18**	**60205542**	**0.3437**	**A**	**C**	**-0.0820149**	**$1.46x10^{-8}$**	**Intronic**	**ZCCHC2**
rs12000047	9	105728627	0.01119	A	G	-0.267122	$5.08x10^{-8}$	-	-
rs12005200	9	105708946	0.01119	G	A	-0.264636	$5.77x10^{-8}$	-	-
rs6502321	17	13855742	0.07297	A	G	0.118734	$1.00x10^{-7}$	-	-
rs4753426	11	92701596	0.4706	G	A	-0.0721806	$4.07x10^{-7}$	-	MTNR1B(-1.192kb)
rs12866867	13	45252171	0.2276	A	G	-0.0793227	$1.34x10^{-6}$	-	-
rs6483212	11	92725240	0.4884	A	G	0.0687848	$1.76x10^{-6}$	-	MTNR1B(+9.292kb)
rs4753073	11	92717475	0.4876	G	A	0.0685863	$1.89x10^{-6}$	-	MTNR1B(+1.527kb)
rs11754507	6	116697130	0.01701	G	A	0.251601	$2.11x10^{-6}$	Intronic	DSE
rs9863500	3	112440948	0.4917	A	G	-0.0694618	$4.07x10^{-6}$	-	LOC101929694(-14.35kb)
rs11259593	10	15320660	0.01042	A	G	0.181618	$5.35x10^{-6}$	Intronic	FAM171A1
rs9998501	4	30403871	0.0199	G	A	-0.157791	$6.46x10^{-6}$	-	-
rs6661803	1	117016646	0.01617	A	G	-0.263019	$7.90x10^{-6}$	-	-
rs7257519	19	57474120	0.4722	A	G	0.0670852	$8.35x10^{-6}$	-	-
rs17041322	4	110982272	0.01534	A	C	0.190737	$8.40x10^{-6}$	-	ELOVL6
rs17126828	14	90969615	0.08638	A	G	-0.116789	$9.69x10^{-6}$	-	-
Joint effect									
rs17070072	**18**	**60205542**	**0.3437**	**A**	**C**	**-0.07339185**	**$1.96x10^{-8}$**	**Intronic**	**ZCCHC2**
rs12000047	**9**	**105728627**	**0.01119**	**A**	**G**	**-0.2313853**	**$2.00x10^{-8}$**	**-**	**-**
rs12005200	**9**	**105708946**	**0.01119**	**G**	**A**	**-0.2282324**	**$2.09x10^{-8}$**	**-**	**-**
rs6502321	17	13855742	0.07297	A	G	0.0521247	$3.26x10^{-7}$	-	-
rs958033	18	1936530	0.1777	G	A	-0.0864242	$1.29x10^{-6}$	-	LINC00540(+18.03kb)
rs8091041	18	1931566	0.1776	G	A	-0.086394	$1.29x10^{-6}$	-	-
rs4409393	8	57214372	0.3665	A	G	-0.0193113	$1.62x10^{-6}$	Intronic	SDR16C5
rs4753426	11	92701596	0.4706	G	A	-0.055935	$2.13x10^{-6}$	-	MTNR1B(-1.192kb)
rs11754507	6	116697130	0.01701	G	A	0.14902	$3.34x10^{-6}$	Intronic	DSE
rs7257519	19	57474120	0.4722	A	G	0.0238733	$4.58x10^{-6}$	-	-
rs11259022	10	14385164	0.1712	A	G	0.0448526	$4.64x10^{-6}$	-	FRMD4A(+12.3kb)

rs6483212	11	92725240	0.4884	A	G	0.0582159	$5.52x10^{-6}$	-	MTNR1B(+9.292kb)
rs4753073	11	92717475	0.4876	G	A	0.0579492	$6.03x10^{-6}$	-	MTNR1B(+1.527kb)
rs12866867	13	45252171	0.2276	A	G	-0.0564297	$7.64x10^{-6}$	-	-
rs10811635	9	21910698	0.1907	G	A	0.07892428	$8.56x10^{-6}$	-	-

Supplementary Table 10. Top ISLE GWEIS results in Generation Scotland ($p<1x10^{-5}$) - Independent SLE as exposure

SNP	CHR	BP	MAF	A1	A2	BETA	*p*-value	Location	Closest gene (<25kb)
GxE effect									
rs2803589	10	88372958	0.1501	G	A	-0.0742497	$1.56x10^{-6}$	-	-
rs11628732	14	58514673	0.1513	G	A	0.0701941	$1.80x10^{-6}$	Intronic	C14orf37
rs3825243	12	69750839	0.04063	A	G	0.109939	$1.98x10^{-6}$	-	LYZ(+2.826kb)|YEATS4(-2.692kb)
rs12677170	8	85446567	0.1169	C	A	-0.074386	$2.09x10^{-6}$	Intronic	RALYL
rs10250565	7	82370821	0.03151	A	G	-0.131706	$2.50x10^{-6}$	-	PCLO(-12.5kb)
rs17548315	6	3114947	0.01824	A	G	0.141617	$2.57x10^{-6}$	Exonic	RIPK1
rs800336	11	2473131	0.2504	G	A	0.0558322	$4.40x10^{-6}$	Intronic	KCNQ1
rs2225291	14	57145062	0.3308	G	A	0.0511159	$4.91x10^{-6}$	-	-
rs8017903	14	57146809	0.3317	A	G	0.0509381	$5.43x10^{-6}$	-	-
rs12432233	14	57141318	0.3481	A	C	0.0497956	$8.12x10^{-6}$	-	-
rs716651	5	155087160	0.4005	A	G	-0.0465724	$9.14x10^{-6}$	-	-
rs1579282	5	155074865	0.3308	A	G	-0.0489933	$9.68x10^{-6}$	-	-
Joint effect									
rs1579282	5	155074865	0.3308	A	G	-0.04726137	$1.94x10^{-7}$	-	-
rs716651	5	155087160	0.4005	A	G	-0.04079837	$6.97x10^{-7}$	-	-
rs17244735	13	73104490	0.04892	G	A	0.1177909	$8.04x10^{-7}$	-	-
rs9879244	3	176614990	0.08416	A	G	-0.0781681	$3.45x10^{-6}$	-	-
rs17548315	6	3114947	0.01824	A	G	0.0764885	$3.89x10^{-6}$	Exonic	RIPK1
rs2242223	5	121761461	0.2384	G	A	-0.0156052	$5.08x10^{-6}$	Intronic	SNCAIP
rs12677170	8	85446567	0.1169	C	A	-0.0295941	$5.27x10^{-6}$	Intronic	RALYL
rs2803589	10	88372958	0.1501	G	A	-0.0324028	$5.69x10^{-6}$	-	-
rs16870827	5	70976503	0.1787	G	A	0.0690271	$5.88x10^{-6}$	-	MCCC2(+21.97kb)
rs2290987	5	121776301	0.2289	G	A	-0.0097764	$9.23x10^{-6}$	Intronic|Intronic	MGC32805|SNCAIP
rs3811891	5	121765801	0.2318	G	A	-0.011288	$9.77x10^{-6}$	Intronic	SNCAIP

Supplementary Table 11. Gene sets enriched by genes prioritized by FUMA from UK Biobank GWEIS

N: The total number of genes in the gene set.

n: The total number of prioritized genes in the gene set.

GWEIS for GxE effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GWAS catalog					
Post bronchodilator FEV1/FVC ratio	196	2	2.02×10^{-5}	1.45×10^{-3}	MDGA2, HSDL2

GWEIS for joint effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
All Canonical Pathways					
biocarta gpcr pathway	34	2	1.80×10^{-6}	2.39×10^{-3}	CREB1, GNGT1
reactome opioid signalling	77	2	2.16×10^{-5}	5.22×10^{-3}	CREB1, GNGT1
reactome neurotransmitter receptor binding and downstream transmission in the postsynaptic cell	133	2	1.10×10^{-4}	5.37×10^{-3}	CREB1, GNGT1
reactome transmission across chemical synapses	181	2	2.74×10^{-4}	6.07×10^{-3}	CREB1, GNGT1
reactome gastrin creb signalling pathway via pkc and mapk	200	2	3.68×10^{-4}	6.78×10^{-3}	CREB1, GNGT1
reactome neuronal system	273	2	9.10×10^{-4}	1.27×10^{-2}	CREB1, GNGT1
GO biological processes					
bone morphogenesis	79	2	2.33×10^{-5}	2.68×10^{-2}	COL13A1, GLG1
bone development	155	2	1.74×10^{-4}	2.79×10^{-2}	COL13A1, GLG1
skeletal system morphogenesis	200	2	3.68×10^{-4}	3.86×10^{-2}	COL13A1, GLG1
neuron projection guidance	204	2	3.90×10^{-4}	3.86×10^{-2}	KLF7, CREB1
GO molecular functions					
transcription factor activity rna polymerase ii distal enhancer sequence specific binding	89	2	3.33×10^{-5}	1.07×10^{-2}	CREB1, PURA
transcription factor activity protein binding	580	3	6.38×10^{-4}	2.55×10^{-2}	KLF7, CREB1, PHF2
transcription coactivator activity	295	2	1.14×10^{-3}	3.77×10^{-2}	KLF7, PHF2
GWAS catalog					
Migraine	76	2	2.08×10^{-5}	3.19×10^{-3}	AGBL1, IMMP2L
Body mass index	335	3	7.80×10^{-5}	5.60×10^{-3}	CREB1, FAM120AOS, FAM120A

Supplementary Table 12. Gene set enriched by genes prioritized by FUMA from Generation Scotland TSLE GWEIS using total SLE as exposure

N: The total number of genes in the gene set.
n: The total number of prioritized genes in the gene set.

TSLE GWEIS for GxE effect

GeneSet	N	n	*p*-value	Adjusted *p*	genes
GO biological processes					
regulation of jun kinase activity	81	2	3.13×10^{-5}	3.40×10^{-2}	AIDA, RIPK1
regulation of jnk cascade	159	2	2.33×10^{-4}	4.03×10^{-2}	AIDA, RIPK1
endoplasmatic reticulum to golgi vesicle mediated transport	165	2	2.60×10^{-4}	4.15×10^{-2}	MIA3, TRAPPC3L
GWAS catalog					
Age-related macular degeneration	64	2	1.54×10^{-5}	5.22×10^{-3}	FRK, COL10A1
Myopia	72	2	2.20×10^{-5}	5.22×10^{-3}	NT5DC1, COL10A1
Urate levels	79	2	2.91×10^{-5}	5.22×10^{-3}	MIA3, FRK

TSLE GWEIS for joint effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GO biological processes					
protein autophosphorylation	192	2	3.26×10^{-4}	4.34×10^{-2}	RIPK1, FRK
GWAS catalog					
Heschl's gyrus morphology	35	2	1.97×10^{-6}	2.12×10^{-3}	KALRN, DSE
Age-related macular degeneration	64	2	1.24×10^{-5}	3.68×10^{-3}	FRK, COL10A1
Myopia	72	2	1.76×10^{-5}	3.80×10^{-3}	NT5DC1, COL10A1

Supplementary Table 13. Gene set enriched by genes prioritized by FUMA from Generation Scotland DSLE GWEIS using dependent SLE as exposure.

N: The total number of genes in the gene set.
n: The total number of prioritized genes in the gene set.

DSLE GWEIS for GxE effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GWAS catalog					
Age-related macular degeneration	64	2	2.89×10^{-6}	8.89×10^{-4}	FRK, COL10A1
Myopia	72	2	4.13×10^{-6}	8.89×10^{-4}	NT5DC1, COL10A1

DSLE GWEIS for joint effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GO biological processes					
sulfur compound biosynthetic process	201	2	1.73×10^{-4}	2.24×10^{-2}	DSE, MTAP
carboxylic acid biosynthetic process	264	2	3.86×10^{-4}	3.45×10^{-2}	DSE, MTAP
organic acid biosynthetic process	264	2	3.86×10^{-4}	3.45×10^{-2}	DSE, MTAP
GWAS catalog					
Age-related macular degeneration	64	2	5.65×10^{-6}	1.34×10^{-3}	FRK, COL10A1
Myopia	72	2	8.06×10^{-6}	1.45×10^{-3}	NT5DC1, COL10A1

Supplementary Table 14. Gene set enriched by genes prioritized by FUMA from Generation Scotland DSLE GWEIS using independent SLE as exposure.

N: The total number of genes in the gene set.
n: The total number of prioritized genes in the gene set.

ISLE GWEIS for GxE effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GO biological processes					
regulation of transporter activity	198	2	2.08×10^{-5}	3.46×10^{-3}	KCNQ1, RIPK1
regulation of transmembrane transport	423	2	2.00×10^{-4}	1.24×10^{-2}	KCNQ1, RIPK1
regulation of ion transport	588	2	5.27×10^{-4}	2.14×10^{-2}	KCNQ1, RIPK1
regulation of secretion	694	2	8.56×10^{-4}	3.03×10^{-2}	KCNQ1, PCLO
regulation of transport	1795	3	9.62×10^{-4}	3.22×10^{-2}	KCNQ1, RIPK1, PCLO
regulation of defense response	755	2	1.09×10^{-3}	3.53×10^{-2}	KCNQ1, RIPK1
cell cell signaling	757	2	1.10×10^{-3}	3.54×10^{-2}	KCNQ1, PCLO
GO molecular functions					
phospholipid binding	355	2	1.19×10^{-4}	1.00×10^{-2}	KCNQ1, PCLO
lipid binding	649	2	7.03×10^{-4}	3.11×10^{-2}	KCNQ1, PCLO
GO cellular components					
membrane microdomain	286	2	6.25×10^{-5}	1.21×10^{-2}	KCNQ1, RIPK1
plasma membrane protein complex	507	2	3.41×10^{-4}	3.31×10^{-2}	KCNQ1, RIPK1
GWAS catalog					
Body mass index	335	2	1.00×10^{-4}	1.19×10^{-2}	KCNQ1, RALYL

ISLE GWEIS for joint effect

GeneSet	N	n	*p*-value	adjusted *p*	genes
GO molecular functions					
ubiquitin like protein ligase binding	263	2	2.46×10^{-5}	1.14×10^{-2}	SNCAIP, RIPK1
identical protein binding	1202	3	7.14×10^{-5}	1.66×10^{-2}	SNCAIP, RIPK1, RALYL

Supplementary Table 15. Cross-cohort prediction.

Phenotype: Name of phenotype predicted. (R) refers to phenotypes using mapping by proxy approach (i.e. where first-degree relatives of individuals with the disease are consider proxy cases and included into the group of cases).
Summary.statistics: Summary statistics used to weighted PRS.
Threshold: P-value threshold used to construct best PRS predictor.
Num_SNP: Number of SNPs included in the model.
PRS.R2: Phenotypic variance explained by the PRS.
Full.R2: Phenotypic variance explained by the full model (including the covariates).
Null.R2: Phenotypic variance explained by the covariates.
Coefficient: Regression coefficient of the model (can provide insight of the direction of effect).
P: P value of the model fit.
Empirical-P: Empirical p-value after permuting 10,000 times P value of the model fit (it should account for multiple testing and over-fitting). In red Empirical-P < 0.05.

Predicting in Generation Scotland:

Phenotype	Summary.statistics	Threshold	Num_SNP	PRS.R2	Full.R2	Null.R2	Coefficient	Standard Error	*p*-value	Empirical-*p*
GHQ	PGC2-MDD GWAS	0.3	42768	0.00779085	0.0326142	0.0248234	0.0479313	0.00763397	3.71×10^{-10}	1.00×10^{-4}
GHQ	UKB GWAS	0.5	64419	0.00564066	0.030464	0.0248234	0.0403361	0.00755848	9.90×10^{-8}	1.00×10^{-4}
GHQ	UKB GWEIS - GxE effect	0.03	7259	0.00153355	0.0263569	0.0248234	0.0211521	0.00761778	0.00551253	0.029897
GHQ	UKB GWEIS - Joint effect	0.4	55348	0.00578933	0.0306127	0.0248234	4.11×10^{-2}	0.00760206	6.72×10^{-8}	1.00×10^{-4}

Predicting in UK Biobank:

Phenotype	Summary.statistics	Threshold	Num_SNP	PRS.R2	Full.R2	Null.R2	Coefficient	Standard Error	*p*-value	Empirical-*p*
PHQ	PGC2-MDD GWAS	0.5	57356	0.0044533	0.0373778	0.0329245	0.0590839	0.00277599	2.69×10^{-100}	1.00×10^{-4}
PHQ	GS GWAS	0.2	34755	0.000403055	0.0333276	0.0329245	0.0177551	0.00276299	1.32×10^{-10}	1.00×10^{-4}
PHQ	GS GWEIS using TSLE - GxE effect	0.001	526	2.60×10^{-5}	0.0329505	0.0329245	-0.00450364	0.00276207	0.102993	0.382062
PHQ	GS GWEIS using DSLE - GxE effect	0.001	522	1.51×10^{-5}	0.0329397	0.0329245	-0.00344122	0.00276271	0.212916	0.642236

PHQ	GS GWEIS using ISLE - GxE effect	0.02	5440	$1.18x10^{-5}$	0.0329363	0.0329245	-0.0030395	0.00276926	0.272387	0.748225
PHQ	GS GWEIS using TSLE - Joint effect	1	94403	0.000322961	0.0332475	0.0329245	0.0158888	$2.76x10^{-3}$	$8.85x10^{-9}$	$1.00x10^{-4}$
PHQ	GS GWEIS using DSLE - Joint effect	0.5	65749	0.000119752	0.0330443	0.0329245	0.00969592	$2.77x10^{-3}$	$4.62x10^{-4}$	0.00429957
PHQ	GS GWEIS using ISLE - Joint effect	1	94329	0.000316573	0.0332411	0.0329245	0.0157459	$2.76x10^{-3}$	$1.24x10^{-8}$	$1.00x10^{-4}$

Supplementary Table 16. Prediction of stress-related traits.

Phenotype: Name of phenotype predicted. (R) refers to phenotypes using mapping by proxy approach (i.e. where first-degree relatives of individuals with the disease are consider proxy cases and included into the group of cases).
Summary.statistics: Summary statistics used to weighted PRS.
Threshold: P-value threshold used to construct best PRS predictor.
Num_SNP: Number of SNPs included in the model.
PRS.R2: Phenotypic variance explained by the PRS.
Full.R2: Phenotypic variance explained by the full model (including the covariates).
Null.R2: Phenotypic variance explained by the covariates.
Coefficient: Regression coefficient of the model (can provide insight of the direction of effect).
P: P value of the model fit.
Empirical-P: Empirical p-value after permuting 10,000 times P value of the model fit (it should account for multiple testing and over-fitting). In red Empirical-P < 0.05.
FDR Empirical-P adjusted: Empirical-P adjested by FDR (300 tests). **In bold red FDR Empirical-P adjusted < 0.05.**

Phenotype	Summary.statistics	Threshold	Num_SNP	PRS.R2	Full.R2	Null.R2	Coefficient	Standard.Error	*p*-value	Empirical-*p*	FDR Empirical.P adjusted
Alzheimer (R)	GS GWAS	0.001	342	0.00069	0.01761	0.01692	0.05362	0.04405	0.22349	0.53805	0.82248
Alzheimer (R)	GS GWEIS using DSLE - GxE effect	0.5	67086	0.00078	0.01770	0.01692	-0.05687	0.04401	0.19630	0.47595	0.79064
Alzheimer (R)	GS GWEIS using DSLE - Joint effect	0.4	57718	0.00126	0.01818	0.01692	-0.07254	0.04436	0.10201	0.30207	0.64729
Alzheimer (R)	GS GWEIS using ISLE - GxE effect	0.001	441	0.00047	0.01739	0.01692	0.04398	0.04368	0.31399	0.67403	0.88618
Alzheimer (R)	GS GWEIS using ISLE - Joint effect	0.001	421	0.00451	0.02143	0.01692	0.13717	0.04423	0.00193	0.00890	0.07852
Alzheimer (R)	GS GWEIS using TSLE - GxE effect	0.001	530	0.00007	0.01699	0.01692	0.01764	0.04407	0.68900	0.98740	0.99737
Alzheimer (R)	GS GWEIS using TSLE - Joint effect	0.001	473	0.00257	0.01949	0.01692	0.10353	0.04421	0.01920	0.07029	0.34997
Alzheimer (R)	UKB GWAS	1	90832	0.00053	0.01745	0.01692	0.04718	0.04426	0.28638	0.74433	0.88963
Alzheimer (R)	UKB GxE effect	0.03	7265	0.00042	0.01734	0.01692	0.04198	0.04437	0.34407	0.86201	0.93359

Alzheimer (R)	UKB GxE Joint effect	1	91064	0.00058	0.01750	0.01692	0.04960	0.04449	0.26497	0.75393	0.89384
Asthma	GS GWAS	0.001	342	0.00136	0.04156	0.04020	-0.07963	0.04772	0.09516	0.26637	0.62923
Asthma	GS GWEIS using DSLE - GxE effect	0.2	35852	0.00043	0.04063	0.04020	0.04363	0.04647	0.34784	0.72433	0.88618
Asthma	GS GWEIS using DSLE - Joint effect	0.05	12712	0.00119	0.04139	0.04020	0.07315	0.04671	0.11730	0.34207	0.69809
Asthma	GS GWEIS using ISLE - GxE effect	0.005	1666	0.00273	0.04293	0.04020	-0.11483	0.04890	0.01886	0.06299	0.34997
Asthma	GS GWEIS using ISLE - Joint effect	1	95515	0.00087	0.04107	0.04020	0.06372	0.04757	0.18038	0.47775	0.79064
Asthma	GS GWEIS using TSLE - GxE effect	0.1	21099	0.00004	0.04024	0.04020	-0.01381	0.04800	0.77363	0.99890	0.99950
Asthma	GS GWEIS using TSLE - Joint effect	0.1	21700	0.00048	0.04068	0.04020	0.04786	0.04831	0.32187	0.71963	0.88618
Asthma	UKB GWAS	0.2	34864	0.00063	0.04083	0.04020	-0.05419	0.04766	0.25545	0.69863	0.88618
Asthma	UKB GxE effect	0.1	19894	0.00077	0.04097	0.04020	0.06028	0.04810	0.21014	0.65753	0.87456
Asthma	UKB GxE Joint effect	0.005	1927	0.00067	0.04087	0.04020	0.05542	0.04742	0.24249	0.72903	0.88618
Asthma (R)	GS GWAS	0.5	66423	0.00204	0.11480	0.11276	0.09352	0.04125	0.02338	0.07879	0.36366
Asthma (R)	GS GWEIS using DSLE - GxE effect	0.005	1907	0.00031	0.11307	0.11276	-0.03599	0.04066	0.37613	0.76452	0.89384
Asthma (R)	GS GWEIS using DSLE - Joint effect	1	95559	0.00116	0.11392	0.11276	0.06855	0.04025	0.08856	0.27707	0.63589
Asthma (R)	GS GWEIS using ISLE - GxE effect	0.001	441	0.00065	0.11342	0.11276	0.05250	0.04085	0.19880	0.48355	0.79271
Asthma (R)	GS GWEIS using ISLE - Joint effect	0.2	35677	0.00153	0.11429	0.11276	0.08069	0.04100	0.04905	0.16628	0.49819
Asthma (R)	GS GWEIS using TSLE - GxE effect	0.005	1836	0.00100	0.11376	0.11276	0.06457	0.04065	0.11222	0.29517	0.64729
Asthma (R)	GS GWEIS using TSLE - Joint effect	0.1	21704	0.00151	0.11427	0.11276	0.08102	0.04150	0.05089	0.16468	0.49819
Asthma (R)	UKB GWAS	0.4	56075	0.00100	0.11376	0.11276	0.06493	0.04088	0.11218	0.38966	0.74458
Asthma (R)	UKB GxE effect	0.1	19901	0.00100	0.11377	0.11276	0.06585	0.04142	0.11184	0.42846	0.76164
Asthma (R)	UKB GxE Joint effect	0.4	55274	0.00064	0.11340	0.11276	0.05209	0.04115	0.20565	0.64194	0.86748
Bowel cancer (R)	GS GWAS	0.04	9529	0.00122	0.01221	0.01099	0.07018	0.04321	0.10433	0.29037	0.64729

Bowel cancer (R)	GS GWEIS using DSLE - GxE effect	0.001	523	0.00167	0.01266	0.01099	-0.08262	0.04367	0.05848	0.16638	0.49819
Bowel cancer (R)	GS GWEIS using DSLE - Joint effect	0.001	503	0.00044	0.01142	0.01099	-0.04205	0.04342	0.33291	0.74273	0.88963
Bowel cancer (R)	GS GWEIS using ISLE - GxE effect	0.005	1666	0.00003	0.01101	0.01099	-0.01025	0.04374	0.81475	0.99950	0.99950
Bowel cancer (R)	GS GWEIS using ISLE - Joint effect	0.5	65995	0.00050	0.01149	0.01099	0.04486	0.04315	0.29854	0.68713	0.88618
Bowel cancer (R)	GS GWEIS using TSLE - GxE effect	0.005	1840	0.00059	0.01157	0.01099	0.04882	0.04340	0.26060	0.59124	0.84063
Bowel cancer (R)	GS GWEIS using TSLE - Joint effect	0.4	57528	0.00074	0.01172	0.01099	0.05449	0.04328	0.20807	0.51985	0.82248
Bowel cancer (R)	UKB GWAS	0.03	8341	0.00102	0.01200	0.01099	0.06476	0.04384	0.13964	0.46365	0.78144
Bowel cancer (R)	UKB GxE effect	0.005	1459	0.00081	0.01179	0.01099	-0.05777	0.04379	0.18714	0.60854	0.85780
Bowel cancer (R)	UKB GxE Joint effect	0.5	63553	0.00092	0.01191	0.01099	-0.06209	0.04406	0.15873	0.54915	0.82248
Breast cancer	GS GWAS	0.02	5289	0.00383	0.20373	0.19991	-0.20547	0.12414	0.09789	0.28087	0.63835
Breast cancer	GS GWEIS using DSLE - GxE effect	0.001	523	0.00525	0.20516	0.19991	0.23555	0.11937	0.04847	0.16488	0.49819
Breast cancer	GS GWEIS using DSLE - Joint effect	0.02	6098	0.00436	0.20427	0.19991	-0.22352	0.12417	0.07183	0.25038	0.61568
Breast cancer	GS GWEIS using ISLE - GxE effect	1	95353	0.00116	0.20107	0.19991	-0.11006	0.11978	0.35817	0.73273	0.88618
Breast cancer	GS GWEIS using ISLE - Joint effect	0.05	12306	0.00515	0.20506	0.19991	-0.23333	0.12080	0.05341	0.18148	0.51852
Breast cancer	GS GWEIS using TSLE - GxE effect	0.001	531	0.00421	0.20411	0.19991	0.20565	0.11644	0.07737	0.22708	0.59201
Breast cancer	GS GWEIS using TSLE - Joint effect	0.001	474	0.00355	0.20346	0.19991	-0.18852	0.11771	0.10925	0.32357	0.66959
Breast cancer	UKB GWAS	0.4	56115	0.00861	0.20852	0.19991	-0.29322	0.11810	0.01303	0.06379	0.34997
Breast cancer	UKB GxE effect	0.03	7258	0.00364	0.20354	0.19991	0.18900	0.11632	0.10418	0.40836	0.75158
Breast cancer	UKB GxE Joint effect	1	91009	0.00160	0.20151	0.19991	-0.12611	0.11707	0.28137	0.77982	0.90002
Breast	GS GWAS	0.001	342	0.00103	0.00974	0.00871	-0.06723	0.04668	0.14982	0.38206	0.73473

cancer (R)											
Breast cancer (R)	GS GWEIS using DSLE - GxE effect	0.005	1912	0.00029	0.00900	0.00871	-0.03599	0.04683	0.44219	0.84052	0.91911
Breast cancer (R)	GS GWEIS using DSLE - Joint effect	0.01	3432	0.00066	0.00937	0.00871	-0.05419	0.04703	0.24924	0.61634	0.85978
Breast cancer (R)	GS GWEIS using ISLE - GxE effect	0.005	1666	0.00071	0.00942	0.00871	0.05532	0.04610	0.23016	0.53275	0.82248
Breast cancer (R)	GS GWEIS using ISLE - Joint effect	0.001	421	0.00106	0.00977	0.00871	-0.06834	0.04664	0.14288	0.40166	0.74843
Breast cancer (R)	GS GWEIS using TSLE - GxE effect	0.2	35575	0.00005	0.00876	0.00871	0.01518	0.04689	0.74621	0.99650	0.99950
Breast cancer (R)	GS GWEIS using TSLE - Joint effect	0.001	473	0.00240	0.01111	0.00871	-0.10277	0.04674	0.02788	0.09939	0.38304
Breast cancer (R)	UKB GWAS	0.001	556	0.00224	0.01094	0.00871	-0.10028	0.04718	0.03356	0.14849	0.49819
Breast cancer (R)	UKB GxE effect	0.03	7257	0.00134	0.01005	0.00871	-0.07736	0.04706	0.10024	0.39416	0.74841
Breast cancer (R)	UKB GxE Joint effect	0.005	1930	0.00078	0.00949	0.00871	-0.05876	0.04678	0.20907	0.65603	0.87456
COPD	GS GWAS	0.005	1543	0.00437	0.11993	0.11556	-0.21163	0.12635	0.09393	0.27457	0.63589
COPD	GS GWEIS using DSLE - GxE effect	0.5	67087	0.00428	0.11984	0.11556	-0.19134	0.09598	0.04621	0.15828	0.49819
COPD	GS GWEIS using DSLE - Joint effect	0.5	66455	0.00590	0.12146	0.11556	-0.23792	0.10671	0.02578	0.11549	0.43308
COPD	GS GWEIS using ISLE - GxE effect	0.005	1669	0.00619	0.12175	0.11556	-0.24883	0.12404	0.04485	0.15399	0.49819
COPD	GS GWEIS using ISLE - Joint effect	0.001	421	0.00470	0.12026	0.11556	-0.21220	0.12191	0.08176	0.25997	0.62394
COPD	GS GWEIS using TSLE - GxE effect	0.3	47860	0.00590	0.12146	0.11556	-0.22583	0.10703	0.03486	0.13109	0.46795
COPD	GS GWEIS using TSLE - Joint effect	0.001	473	0.00496	0.12053	0.11556	-0.21840	0.12205	0.07354	0.23688	0.59201
COPD	UKB GWAS	0.005	2004	0.00548	0.12104	0.11556	0.22961	0.12200	0.05982	0.23618	0.59201
COPD	UKB GxE effect	0.02	5183	0.00080	0.11636	0.11556	-0.08756	0.12170	0.47188	0.96220	0.99279
COPD	UKB GxE Joint effect	0.01	3375	0.00762	0.12318	0.11556	0.26858	0.12104	0.02649	0.13809	0.47616
COPD (R)	GS GWAS	0.5	66477	0.00505	0.03944	0.03439	0.15006	0.04652	0.00126	0.00560	0.05249
COPD (R)	GS GWEIS using	0.1	21401	0.00045	0.03484	0.03439	-0.04557	0.04744	0.33669	0.70383	0.88618

	DSLE - GxE effect										
COPD (R)	GS GWEIS using DSLE - Joint effect	0.01	3433	0.00186	0.03625	0.03439	0.09176	0.04691	0.05047	0.16908	0.49819
COPD (R)	GS GWEIS using ISLE - GxE effect	1	95337	0.00044	0.03483	0.03439	-0.04570	0.04865	0.34761	0.72663	0.88618
COPD (R)	GS GWEIS using ISLE - Joint effect	0.005	1725	0.00450	0.03889	0.03439	0.14340	0.04726	0.00241	0.01010	0.08416
COPD (R)	GS GWEIS using TSLE - GxE effect	0.001	531	0.00356	0.03795	0.03439	-0.12862	0.04790	0.00725	0.02780	0.19393
COPD (R)	**GS GWEIS using TSLE - Joint effect**	**0.01**	**3322**	**0.00709**	**0.04148**	**0.03439**	**0.17978**	**0.04721**	**0.00014**	**0.00080**	**0.01000**
COPD (R)	UKB GWAS	0.001	555	0.00053	0.03492	0.03439	-0.04974	0.04787	0.29872	0.77512	0.90002
COPD (R)	UKB GxE effect	0.2	33863	0.00222	0.03661	0.03439	0.10221	0.04810	0.03361	0.16498	0.49819
COPD (R)	UKB GxE Joint effect	1	91066	0.00036	0.03475	0.03439	0.04107	0.04779	0.39012	0.90861	0.95980
Depression	**GS GWAS**	**0.1**	**20325**	**0.01066**	**0.07495**	**0.06428**	**0.23175**	**0.05000**	**0.00000**	**0.00010**	**0.00200**
Depression	GS GWEIS using DSLE - GxE effect	0.001	524	0.00037	0.06465	0.06428	0.04298	0.05042	0.39393	0.78102	0.90002
Depression	**GS GWEIS using DSLE - Joint effect**	**0.005**	**1912**	**0.00689**	**0.07117**	**0.06428**	**0.18269**	**0.04938**	**0.00022**	**0.00080**	**0.01000**
Depression	GS GWEIS using ISLE - GxE effect	0.005	1667	0.00393	0.06821	0.06428	-0.14306	0.05189	0.00584	0.02180	0.15856
Depression	**GS GWEIS using ISLE - Joint effect**	**0.2**	**35692**	**0.00821**	**0.07249**	**0.06428**	**0.20150**	**0.04956**	**0.00005**	**0.00030**	**0.00474**
Depression	GS GWEIS using TSLE - GxE effect	0.001	531	0.00113	0.06541	0.06428	-0.07536	0.05062	0.13657	0.35497	0.71952
Depression	**GS GWEIS using TSLE - Joint effect**	**0.2**	**36078**	**0.00781**	**0.07209**	**0.06428**	**0.20093**	**0.05070**	**0.00007**	**0.00040**	**0.00600**
Depression	**UKB GWAS**	**0.01**	**3542**	**0.00629**	**0.07057**	**0.06428**	**0.17820**	**0.05077**	**0.00045**	**0.00280**	**0.02896**
Depression	UKB GxE effect	0.04	9348	0.00192	0.06620	0.06428	0.09922	0.05117	0.05249	0.23418	0.59201
Depression	UKB GxE Joint effect	0.05	12391	0.00284	0.06712	0.06428	0.12184	0.05165	0.01833	0.09939	0.38304
Depression (R)	GS GWAS	0.1	20324	0.00169	0.08024	0.07855	0.08651	0.04372	0.04787	0.15169	0.49819
Depression (R)	GS GWEIS using DSLE - GxE effect	0.02	5955	0.00055	0.07910	0.07855	-0.04835	0.04306	0.26151	0.57324	0.82284
Depression (R)	GS GWEIS using DSLE - Joint effect	0.005	1912	0.00091	0.07946	0.07855	0.06189	0.04271	0.14733	0.41386	0.75706

Depression (R)	**GS GWEIS using ISLE - GxE effect**	**0.005**	**1663**	**0.00610**	**0.08466**	**0.07855**	**-0.16738**	**0.04508**	**0.00020**	**0.00070**	**0.00954**
Depression (R)	GS GWEIS using ISLE - Joint effect	1	95537	0.00112	0.07967	0.07855	0.07020	0.04355	0.10697	0.31597	0.66754
Depression (R)	GS GWEIS using TSLE - GxE effect	0.1	21124	0.00118	0.07973	0.07855	-0.07325	0.04489	0.10272	0.27767	0.63589
Depression (R)	GS GWEIS using TSLE - Joint effect	0.005	1846	0.00222	0.08077	0.07855	0.09916	0.04375	0.02343	0.08569	0.36503
Depression (R)	UKB GWAS	0.005	2005	0.00307	0.08162	0.07855	0.11570	0.04350	0.00782	0.03590	0.23040
Depression (R)	UKB GxE effect	0.005	1457	0.00136	0.07992	0.07855	-0.07789	0.04394	0.07629	0.31407	0.66754
Depression (R)	UKB GxE Joint effect	0.01	3372	0.00027	0.07882	0.07855	0.03434	0.04364	0.43137	0.93601	0.97501
Diabetes	GS GWAS	0.04	9516	0.00130	0.08072	0.07942	0.09759	0.07736	0.20714	0.50595	0.82046
Diabetes	GS GWEIS using DSLE - GxE effect	0.005	1908	0.00070	0.08012	0.07942	-0.07403	0.08042	0.35729	0.73443	0.88618
Diabetes	GS GWEIS using DSLE - Joint effect	0.001	502	0.00164	0.08106	0.07942	0.10934	0.07778	0.15978	0.44366	0.77341
Diabetes	GS GWEIS using ISLE - GxE effect	0.02	5484	0.00129	0.08071	0.07942	-0.09902	0.08012	0.21649	0.51155	0.82248
Diabetes	GS GWEIS using ISLE - Joint effect	1	95519	0.00086	0.08028	0.07942	0.07923	0.07726	0.30511	0.70523	0.88618
Diabetes	GS GWEIS using TSLE - GxE effect	0.005	1837	0.00021	0.07963	0.07942	0.03889	0.07771	0.61677	0.96740	0.99391
Diabetes	GS GWEIS using TSLE - Joint effect	0.2	36074	0.00125	0.08067	0.07942	0.09542	0.07680	0.21404	0.54195	0.82248
Diabetes	UKB GWAS	0.001	556	0.00613	0.08555	0.07942	0.21643	0.07996	0.00680	0.03610	0.23040
Diabetes	UKB GxE effect	0.005	1458	0.00141	0.08083	0.07942	0.10012	0.07690	0.19294	0.63014	0.86475
Diabetes	UKB GxE Joint effect	0.001	468	0.00159	0.08101	0.07942	0.10829	0.07831	0.16676	0.57154	0.82284
Diabetes (R)	GS GWAS	1	95406	0.00018	0.02840	0.02822	0.02508	0.03695	0.49724	0.89401	0.95108
Diabetes (R)	GS GWEIS using DSLE - GxE effect	0.005	1909	0.00052	0.02874	0.02822	-0.04229	0.03708	0.25401	0.58704	0.83863
Diabetes (R)	GS GWEIS using DSLE - Joint effect	0.05	12705	0.00009	0.02831	0.02822	-0.01748	0.03697	0.63636	0.98440	0.99737
Diabetes (R)	GS GWEIS using ISLE - GxE effect	0.005	1667	0.00053	0.02875	0.02822	-0.04278	0.03727	0.25106	0.56544	0.82284

Diabetes (R)	GS GWEIS using ISLE - Joint effect	0.1	21221	0.00022	0.02843	0.02822	0.02715	0.03692	0.46212	0.88451	0.94432
Diabetes (R)	GS GWEIS using TSLE - GxE effect	0.001	530	0.00087	0.02909	0.02822	-0.05445	0.03704	0.14151	0.37126	0.72324
Diabetes (R)	GS GWEIS using TSLE - Joint effect	0.005	1843	0.00076	0.02898	0.02822	0.05089	0.03695	0.16842	0.45376	0.77345
Diabetes (R)	UKB GWAS	1	90858	0.00116	0.02937	0.02822	0.06290	0.03706	0.08964	0.32527	0.66959
Diabetes (R)	UKB GxE effect	0.02	5172	0.00317	0.03139	0.02822	-0.10539	0.03748	0.00493	0.03000	0.20453
Diabetes (R)	UKB GxE Joint effect	0.01	3377	0.00065	0.02887	0.02822	0.04749	0.03715	0.20120	0.63204	0.86475
Extraversion	**GS GWAS**	**0.4**	**57527**	**0.00365**	**0.02915**	**0.02550**	**-0.21489**	**0.06011**	**0.00036**	**0.00130**	**0.01500**
Extraversion	GS GWEIS using DSLE - GxE effect	0.5	67060	0.00032	0.02582	0.02550	0.06396	0.06040	0.28969	0.63524	0.86475
Extraversion	GS GWEIS using DSLE - Joint effect	0.4	57695	0.00152	0.02702	0.02550	-0.13873	0.06028	0.02142	0.08219	0.36393
Extraversion	GS GWEIS using ISLE - GxE effect	0.03	7742	0.00089	0.02639	0.02550	0.10652	0.06035	0.07765	0.22118	0.58720
Extraversion	**GS GWEIS using ISLE - Joint effect**	**0.5**	**65950**	**0.00339**	**0.02889**	**0.02550**	**-0.20696**	**0.06013**	**0.00058**	**0.00340**	**0.03290**
Extraversion	GS GWEIS using TSLE - GxE effect	0.2	35569	0.00051	0.02601	0.02550	0.08034	0.06044	0.18381	0.44656	0.77341
Extraversion	**GS GWEIS using TSLE - Joint effect**	**1**	**95501**	**0.00425**	**0.02975**	**0.02550**	**-0.23166**	**0.06009**	**0.00012**	**0.00070**	**0.00954**
Extraversion	UKB GWAS	0.2	34868	0.00028	0.02578	0.02550	0.05994	0.06079	0.32414	0.81072	0.91092
Extraversion	UKB GxE effect	0.03	7260	0.00095	0.02645	0.02550	0.11081	0.06103	0.06951	0.29717	0.64729
Extraversion	UKB GxE Joint effect	0.001	468	0.00102	0.02652	0.02550	0.11517	0.06112	0.05960	0.25577	0.62384
Heart disease	GS GWAS	1	95409	0.00045	0.16287	0.16242	0.05982	0.07633	0.43325	0.82712	0.91784
Heart disease	GS GWEIS using DSLE - GxE effect	0.02	5949	0.00181	0.16423	0.16242	-0.12700	0.07947	0.11005	0.30177	0.64729
Heart disease	GS GWEIS using DSLE - Joint effect	0.04	10654	0.00044	0.16286	0.16242	0.06146	0.07922	0.43785	0.86571	0.93422
Heart disease	GS GWEIS using ISLE - GxE effect	0.005	1668	0.00416	0.16658	0.16242	-0.18296	0.07779	0.01868	0.07019	0.34997
Heart disease	GS GWEIS using ISLE - Joint effect	0.005	1727	0.00059	0.16302	0.16242	-0.06737	0.07554	0.37248	0.79052	0.90002
Heart	GS GWEIS using	0.01	3311	0.00229	0.16471	0.16242	-0.13552	0.07678	0.07756	0.23298	0.59201

disease	TSLE - GxE effect										
Heart disease	GS GWEIS using TSLE - Joint effect	0.04	10597	0.00064	0.16306	0.16242	0.07048	0.07535	0.34959	0.75293	0.89384
Heart disease	UKB GWAS	0.001	555	0.00134	0.16376	0.16242	-0.10223	0.07614	0.17938	0.55465	0.82248
Heart disease	UKB GxE effect	0.1	19896	0.00450	0.16692	0.16242	0.18467	0.07512	0.01396	0.07299	0.34997
Heart disease	UKB GxE Joint effect	0.05	12389	0.00349	0.16591	0.16242	0.16191	0.07480	0.03041	0.14569	0.49665
Heart disease (R)	GS GWAS	0.001	342	0.00029	0.02523	0.02494	0.03107	0.03699	0.40088	0.79502	0.90002
Heart disease (R)	GS GWEIS using DSLE - GxE effect	1	95437	0.00169	0.02663	0.02494	0.07721	0.03869	0.04599	0.13789	0.47616
Heart disease (R)	GS GWEIS using DSLE - Joint effect	1	95596	0.00162	0.02656	0.02494	0.07508	0.03806	0.04854	0.16548	0.49819
Heart disease (R)	GS GWEIS using ISLE - GxE effect	0.02	5482	0.00054	0.02549	0.02494	-0.04277	0.03686	0.24594	0.56134	0.82284
Heart disease (R)	GS GWEIS using ISLE - Joint effect	0.001	421	0.00133	0.02627	0.02494	0.06723	0.03706	0.06969	0.21838	0.58494
Heart disease (R)	GS GWEIS using TSLE - GxE effect	0.001	529	0.00008	0.02503	0.02494	0.01684	0.03710	0.64991	0.97860	0.99519
Heart disease (R)	GS GWEIS using TSLE - Joint effect	0.001	473	0.00213	0.02708	0.02494	0.08510	0.03707	0.02169	0.07269	0.34997
Heart disease (R)	UKB GWAS	0.01	3552	0.00200	0.02694	0.02494	0.08342	0.03754	0.02628	0.12359	0.44670
Heart disease (R)	UKB GxE effect	0.4	55632	0.00038	0.02533	0.02494	-0.03690	0.03788	0.33007	0.84252	0.91911
Heart disease (R)	UKB GxE Joint effect	0.1	21009	0.00178	0.02673	0.02494	0.07885	0.03755	0.03575	0.17618	0.50822
High BP	GS GWAS	0.4	57528	0.00129	0.16093	0.15964	0.07970	0.04470	0.07456	0.21638	0.58481
High BP	GS GWEIS using DSLE - GxE effect	0.005	1912	0.00054	0.16017	0.15964	-0.05275	0.04600	0.25146	0.57294	0.82284
High BP	GS GWEIS using DSLE - Joint effect	0.2	36247	0.00053	0.16017	0.15964	0.05341	0.04652	0.25093	0.61904	0.85978
High BP	GS GWEIS using ISLE - GxE effect	0.01	2989	0.00017	0.15981	0.15964	0.02878	0.04403	0.51330	0.89811	0.95206
High BP	GS GWEIS using ISLE - Joint effect	1	95533	0.00079	0.16042	0.15964	0.06217	0.04464	0.16374	0.45116	0.77341

High BP	GS GWEIS using TSLE - GxE effect	0.02	5887	0.00011	0.15974	0.15964	-0.02284	0.04493	0.61129	0.96300	0.99279
High BP	GS GWEIS using TSLE - Joint effect	0.001	473	0.00080	0.16043	0.15964	0.06231	0.04457	0.16212	0.42906	0.76164
High BP	UKB GWAS	0.02	6037	0.00052	0.16016	0.15964	0.04988	0.04422	0.25934	0.70373	0.88618
High BP	UKB GxE effect	0.1	19873	0.00081	0.16045	0.15964	0.06368	0.04514	0.15832	0.53995	0.82248
High BP	UKB GxE Joint effect	0.005	1927	0.00054	0.16018	0.15964	0.05153	0.04482	0.25030	0.73553	0.88618
High BP (R)	GS GWAS	0.005	1543	0.00072	0.07199	0.07126	0.04951	0.03576	0.16624	0.41986	0.75963
High BP (R)	GS GWEIS using DSLE - GxE effect	0.03	8308	0.00053	0.07179	0.07126	-0.04279	0.03635	0.23905	0.55185	0.82248
High BP (R)	GS GWEIS using DSLE - Joint effect	0.04	10658	0.00012	0.07139	0.07126	-0.02059	0.03584	0.56568	0.95141	0.98762
High BP (R)	GS GWEIS using ISLE - GxE effect	0.01	2988	0.00075	0.07202	0.07126	0.05061	0.03589	0.15841	0.40086	0.74843
High BP (R)	GS GWEIS using ISLE - Joint effect	0.01	3217	0.00061	0.07187	0.07126	0.04549	0.03575	0.20318	0.53305	0.82248
High BP (R)	GS GWEIS using TSLE - GxE effect	0.005	1838	0.00022	0.07148	0.07126	0.02730	0.03579	0.44563	0.85012	0.92404
High BP (R)	GS GWEIS using TSLE - Joint effect	0.02	5965	0.00038	0.07164	0.07126	0.03571	0.03570	0.31710	0.70453	0.88618
High BP (R)	UKB GWAS	0.01	3547	0.00161	0.07288	0.07126	0.07503	0.03626	0.03852	0.15918	0.49819
High BP (R)	UKB GxE effect	0.1	19904	0.00085	0.07212	0.07126	0.05481	0.03643	0.13241	0.47965	0.79064
High BP (R)	UKB GxE Joint effect	0.005	1927	0.00100	0.07226	0.07126	0.05882	0.03614	0.10359	0.39866	0.74843
Hip fracture (R)	**GS GWAS**	**0.001**	**342**	**0.01296**	**0.02939**	**0.01643**	**0.25448**	**0.05269**	**0.00000**	**0.00010**	**0.00200**
Hip fracture (R)	GS GWEIS using DSLE - GxE effect	0.005	1910	0.00120	0.01764	0.01643	0.07632	0.05160	0.13909	0.35906	0.72295
Hip fracture (R)	**GS GWEIS using DSLE - Joint effect**	**0.005**	**1909**	**0.00687**	**0.02331**	**0.01643**	**0.18154**	**0.05115**	**0.00039**	**0.00300**	**0.03000**
Hip fracture (R)	GS GWEIS using ISLE - GxE effect	0.005	1668	0.00042	0.01686	0.01643	-0.04649	0.05342	0.38413	0.76112	0.89384
Hip fracture (R)	**GS GWEIS using ISLE - Joint effect**	**0.005**	**1726**	**0.01072**	**0.02716**	**0.01643**	**0.22966**	**0.05200**	**0.00001**	**0.00010**	**0.00200**
Hip fracture (R)	GS GWEIS using TSLE - GxE effect	0.001	531	0.00187	0.01830	0.01643	0.09615	0.05227	0.06583	0.19788	0.55480
Hip fracture	**GS GWEIS using**	**0.01**	**3324**	**0.00705**	**0.02349**	**0.01643**	**0.18705**	**0.05220**	**0.00034**	**0.00110**	**0.01320**

(R)	**TSLE - Joint effect**										
Hip fracture (R)	UKB GWAS	0.01	3549	0.00033	0.01677	0.01643	-0.04101	0.05304	0.43936	0.92081	0.96927
Hip fracture (R)	UKB GxE effect	1	91126	0.00083	0.01727	0.01643	0.06597	0.05385	0.22050	0.68233	0.88618
Hip fracture (R)	UKB GxE Joint effect	0.05	12395	0.00078	0.01721	0.01643	-0.06314	0.05327	0.23591	0.70623	0.88618
Lung cancer (R)	GS GWAS	0.4	57548	0.00139	0.02597	0.02458	0.07384	0.04143	0.07473	0.21448	0.58481
Lung cancer (R)	GS GWEIS using DSLE - GxE effect	0.005	1914	0.00015	0.02473	0.02458	-0.02430	0.04102	0.55358	0.93211	0.97433
Lung cancer (R)	GS GWEIS using DSLE - Joint effect	0.3	47876	0.00154	0.02612	0.02458	0.07685	0.04107	0.06127	0.20388	0.56633
Lung cancer (R)	GS GWEIS using ISLE - GxE effect	0.005	1666	0.00060	0.02518	0.02458	-0.04843	0.04160	0.24426	0.55614	0.82248
Lung cancer (R)	GS GWEIS using ISLE - Joint effect	0.1	21248	0.00164	0.02622	0.02458	0.07990	0.04122	0.05258	0.17598	0.50822
Lung cancer (R)	GS GWEIS using TSLE - GxE effect	0.001	531	0.00046	0.02504	0.02458	-0.04207	0.04129	0.30823	0.65883	0.87456
Lung cancer (R)	GS GWEIS using TSLE - Joint effect	0.005	1844	0.00232	0.02690	0.02458	0.09536	0.04144	0.02139	0.07349	0.34997
Lung cancer (R)	UKB GWAS	0.005	2003	0.00052	0.02510	0.02458	0.04522	0.04134	0.27405	0.72663	0.88618
Lung cancer (R)	UKB GxE effect	0.04	9351	0.00223	0.02681	0.02458	0.09440	0.04184	0.02406	0.11749	0.43514
Lung cancer (R)	UKB GxE Joint effect	0.001	469	0.00105	0.02563	0.02458	0.06419	0.04145	0.12144	0.45945	0.77874
Mood disorder	**GS GWAS**	**0.04**	**9515**	**0.01654**	**0.09079**	**0.07425**	**0.41908**	**0.06504**	**0.00000**	**0.00010**	**0.00200**
Mood disorder	GS GWEIS using DSLE - GxE effect	0.005	1913	0.00007	0.07433	0.07425	0.02785	0.06565	0.67145	0.97740	0.99519
Mood disorder	**GS GWEIS using DSLE - Joint effect**	**0.02**	**6081**	**0.00925**	**0.08350**	**0.07425**	**0.31487**	**0.06559**	**0.00000**	**0.00020**	**0.00333**
Mood disorder	GS GWEIS using ISLE - GxE effect	0.001	439	0.00187	0.07612	0.07425	0.14099	0.06565	0.03186	0.09309	0.37476
Mood disorder	**GS GWEIS using ISLE - Joint effect**	**0.04**	**10241**	**0.01447**	**0.08873**	**0.07425**	**0.39236**	**0.06516**	**0.00000**	**0.00010**	**0.00200**
Mood disorder	GS GWEIS using TSLE - GxE effect	0.005	1832	0.00039	0.07464	0.07425	0.06430	0.06568	0.32764	0.67813	0.88618

Mood disorder	**GS GWEIS using TSLE - Joint effect**	**0.1**	**21683**	**0.01276**	**0.08702**	**0.07425**	**0.36820**	**0.06517**	**0.00000**	**0.00010**	**0.00200**
Mood disorder	UKB GWAS	0.03	8329	0.00226	0.07652	0.07425	0.15605	0.06596	0.01808	0.08629	0.36503
Mood disorder	UKB GxE effect	0.1	19903	0.00067	0.07492	0.07425	0.08583	0.06664	0.19785	0.63284	0.86475
Mood disorder	UKB GxE Joint effect	0.1	20988	0.00281	0.07706	0.07425	0.17425	0.06607	0.00841	0.04900	0.29397
Neuroticism	**GS GWAS**	**1**	**95359**	**0.03538**	**0.09100**	**0.05562**	**0.60308**	**0.05244**	**0.00000**	**0.00010**	**0.00200**
Neuroticism	GS GWEIS using DSLE - GxE effect	0.3	47865	0.00137	0.05698	0.05562	-0.11900	0.05360	0.02649	0.09069	0.37270
Neuroticism	**GS GWEIS using DSLE - Joint effect**	**0.1**	**21776**	**0.01876**	**0.07437**	**0.05562**	**0.44034**	**0.05307**	**0.00000**	**0.00010**	**0.00200**
Neuroticism	GS GWEIS using ISLE - GxE effect	0.02	5485	0.00046	0.05608	0.05562	-0.06923	0.05362	0.19670	0.46815	0.78461
Neuroticism	**GS GWEIS using ISLE - Joint effect**	**1**	**95499**	**0.03105**	**0.08666**	**0.05562**	**0.56490**	**0.05257**	**0.00000**	**0.00010**	**0.00200**
Neuroticism	GS GWEIS using TSLE - GxE effect	0.3	47828	0.00129	0.05691	0.05562	-0.11563	0.05363	0.03114	0.10259	0.38958
Neuroticism	**GS GWEIS using TSLE - Joint effect**	**1**	**95507**	**0.03155**	**0.08716**	**0.05562**	**0.56936**	**0.05255**	**0.00000**	**0.00010**	**0.00200**
Neuroticism	**UKB GWAS**	**1**	**90784**	**0.00557**	**0.06118**	**0.05562**	**0.24077**	**0.05364**	**0.00001**	**0.00020**	**0.00333**
Neuroticism	UKB GxE effect	0.001	327	0.00030	0.05592	0.05562	-0.05628	0.05377	0.29534	0.79282	0.90002
Neuroticism	**UKB GxE Joint effect**	**0.2**	**34852**	**0.00370**	**0.05932**	**0.05562**	**0.19771**	**0.05407**	**0.00026**	**0.00160**	**0.01714**
Osteo arthritis	GS GWAS	0.02	5285	0.00130	0.19279	0.19149	0.09234	0.05729	0.10702	0.28907	0.64729
Osteo arthritis	GS GWEIS using DSLE - GxE effect	0.01	3377	0.00043	0.19192	0.19149	-0.05626	0.06103	0.35665	0.73293	0.88618
Osteo arthritis	GS GWEIS using DSLE - Joint effect	0.04	10656	0.00041	0.19190	0.19149	0.05412	0.05966	0.36433	0.79272	0.90002
Osteo arthritis	GS GWEIS using ISLE - GxE effect	0.05	11826	0.00040	0.19188	0.19149	-0.05124	0.05834	0.37984	0.76092	0.89384
Osteo arthritis	GS GWEIS using ISLE - Joint effect	0.005	1726	0.00078	0.19226	0.19149	0.07017	0.05653	0.21452	0.54545	0.82248
Osteo arthritis	GS GWEIS using TSLE - GxE effect	0.005	1838	0.00018	0.19167	0.19149	0.03402	0.05708	0.55122	0.92571	0.97102
Osteo	GS GWEIS using	0.02	5969	0.00134	0.19283	0.19149	0.09344	0.05716	0.10211	0.29917	0.64729

arthritis	TSLE - Joint effect										
Osteo arthritis	UKB GWAS	0.02	6030	0.00044	0.19193	0.19149	-0.05180	0.05542	0.34995	0.83782	0.91911
Osteo arthritis	UKB GxE effect	0.01	2832	0.00143	0.19291	0.19149	-0.09528	0.05687	0.09386	0.36776	0.72324
Osteo arthritis	UKB GxE Joint effect	0.5	63512	0.00057	0.19206	0.19149	-0.05992	0.05642	0.28816	0.78582	0.90002
Osteo arthritis (R)	GS GWAS	0.03	7442	0.00378	0.05821	0.05443	0.11930	0.03922	0.00235	0.01050	0.08513
Osteo arthritis (R)	GS GWEIS using DSLE - GxE effect	0.02	5954	0.00020	0.05462	0.05443	-0.02719	0.03906	0.48630	0.87621	0.93880
Osteo arthritis (R)	GS GWEIS using DSLE - Joint effect	0.02	6102	0.00386	0.05829	0.05443	0.12053	0.03938	0.00221	0.00980	0.08399
Osteo arthritis (R)	GS GWEIS using ISLE - GxE effect	0.005	1666	0.00215	0.05658	0.05443	-0.09118	0.03994	0.02243	0.07179	0.34997
Osteo arthritis (R)	GS GWEIS using ISLE - Joint effect	0.03	8133	0.00320	0.05763	0.05443	0.10962	0.03911	0.00506	0.02220	0.15856
Osteo arthritis (R)	GS GWEIS using TSLE - GxE effect	0.04	10296	0.00051	0.05494	0.05443	-0.04418	0.03969	0.26565	0.60904	0.85780
Osteo arthritis (R)	**GS GWEIS using TSLE - Joint effect**	**0.02**	**5967**	**0.00515**	**0.05957**	**0.05443**	**0.13991**	**0.03944**	**0.00039**	**0.00160**	**0.01714**
Osteo arthritis (R)	UKB GWAS	1	90783	0.00110	0.05552	0.05443	0.06451	0.03939	0.10148	0.36476	0.72324
Osteo arthritis (R)	UKB GxE effect	0.02	5174	0.00236	0.05678	0.05443	-0.09514	0.03962	0.01632	0.08639	0.36503
Osteo arthritis (R)	UKB GxE Joint effect	0.02	5946	0.00041	0.05484	0.05443	-0.03976	0.03972	0.31678	0.83532	0.91911
Parkinson (R)	GS GWAS	0.05	11504	0.00364	0.02723	0.02359	0.14927	0.06623	0.02422	0.08249	0.36393
Parkinson (R)	GS GWEIS using DSLE - GxE effect	0.02	5951	0.00044	0.02403	0.02359	-0.05306	0.06815	0.43616	0.82912	0.91784
Parkinson (R)	GS GWEIS using DSLE - Joint effect	0.005	1909	0.00481	0.02839	0.02359	0.17088	0.06576	0.00936	0.04270	0.26140
Parkinson (R)	GS GWEIS using ISLE - GxE effect	0.005	1667	0.00102	0.02461	0.02359	-0.08202	0.06994	0.24096	0.55654	0.82248
Parkinson (R)	GS GWEIS using ISLE - Joint effect	0.3	47486	0.00388	0.02747	0.02359	0.15239	0.06513	0.01928	0.07659	0.35903
Parkinson (R)	GS GWEIS using TSLE - GxE effect	0.001	531	0.00066	0.02424	0.02359	0.06482	0.06850	0.34398	0.71823	0.88618

Parkinson (R)	GS GWEIS using TSLE - Joint effect	0.03	8416	0.00362	0.02720	0.02359	0.14887	0.06654	0.02527	0.09369	0.37476
Parkinson (R)	UKB GWAS	0.3	46449	0.00389	0.02748	0.02359	-0.15737	0.06852	0.02164	0.09959	0.38304
Parkinson (R)	UKB GxE effect	0.03	7263	0.00152	0.02511	0.02359	-0.09921	0.06916	0.15142	0.52915	0.82248
Parkinson (R)	UKB GxE Joint effect	0.001	469	0.00449	0.02808	0.02359	0.16990	0.06888	0.01364	0.07279	0.34997
Prostate cancer (R)	GS GWAS	0.005	1545	0.00207	0.01276	0.01069	0.10574	0.05782	0.06744	0.19708	0.55480
Prostate cancer (R)	GS GWEIS using DSLE - GxE effect	0.2	35880	0.00031	0.01100	0.01069	-0.04033	0.05697	0.47893	0.87221	0.93786
Prostate cancer (R)	GS GWEIS using DSLE - Joint effect	0.005	1911	0.00049	0.01118	0.01069	0.05134	0.05786	0.37495	0.81052	0.91092
Prostate cancer (R)	GS GWEIS using ISLE - GxE effect	0.001	440	0.00187	0.01256	0.01069	-0.10270	0.05948	0.08425	0.23348	0.59201
Prostate cancer (R)	GS GWEIS using ISLE - Joint effect	0.01	3217	0.00266	0.01335	0.01069	0.11941	0.05754	0.03795	0.13259	0.46795
Prostate cancer (R)	GS GWEIS using TSLE - GxE effect	1	95499	0.00231	0.01299	0.01069	-0.11491	0.05974	0.05443	0.16938	0.49819
Prostate cancer (R)	GS GWEIS using TSLE - Joint effect	0.5	66141	0.00194	0.01263	0.01069	0.10204	0.05736	0.07524	0.23408	0.59201
Prostate cancer (R)	UKB GWAS	0.03	8341	0.00122	0.01191	0.01069	0.08256	0.05913	0.16260	0.52415	0.82248
Prostate cancer (R)	UKB GxE effect	0.2	33862	0.00364	0.01433	0.01069	-0.14411	0.05973	0.01584	0.08099	0.36393
Prostate cancer (R)	UKB GxE Joint effect	0.4	55274	0.00107	0.01176	0.01069	-0.07743	0.05911	0.19021	0.61904	0.85978
Rheumatoid arthritis	GS GWAS	0.02	5288	0.00536	0.09065	0.08529	0.21537	0.10336	0.03719	0.12099	0.44264
Rheumatoid arthritis	GS GWEIS using DSLE - GxE effect	0.01	3375	0.00257	0.08786	0.08529	0.14970	0.10195	0.14202	0.36966	0.72324
Rheumatoid arthritis	GS GWEIS using DSLE - Joint effect	0.005	1910	0.00345	0.08874	0.08529	0.17432	0.10465	0.09575	0.29927	0.64729
Rheumatoid arthritis	GS GWEIS using ISLE - GxE effect	0.04	9860	0.00115	0.08643	0.08529	-0.10459	0.11222	0.35131	0.72323	0.88618
Rheumatoid arthritis	GS GWEIS using ISLE - Joint effect	0.04	10264	0.00221	0.08750	0.08529	0.13956	0.10520	0.18463	0.49525	0.80747
Rheumatoid arthritis	GS GWEIS using TSLE - GxE effect	0.02	5887	0.00082	0.08611	0.08529	0.08315	0.10226	0.41613	0.81622	0.91368

Rheumatoid arthritis	GS GWEIS using TSLE - Joint effect	0.04	10596	0.00263	0.08792	0.08529	0.15289	0.10546	0.14715	0.40566	0.75122
Rheumatoid arthritis	UKB GWAS	1	90845	0.00497	0.09026	0.08529	0.20883	0.10711	0.05121	0.20968	0.57710
Rheumatoid arthritis	UKB GxE effect	0.5	64076	0.00179	0.08708	0.08529	0.12703	0.10847	0.24158	0.72053	0.88618
Rheumatoid arthritis	UKB GxE Joint effect	0.05	12401	0.00267	0.08796	0.08529	0.15451	0.10801	0.15258	0.52905	0.82248
Rheumatoid arthritis (R)	GS GWAS	0.02	5288	0.00239	0.03987	0.03749	0.09792	0.04192	0.01948	0.06899	0.34997
Rheumatoid arthritis (R)	GS GWEIS using DSLE - GxE effect	0.005	1912	0.00078	0.03826	0.03749	0.05557	0.04180	0.18379	0.44906	0.77341
Rheumatoid arthritis (R)	GS GWEIS using DSLE - Joint effect	0.04	10655	0.00288	0.04037	0.03749	0.10647	0.04161	0.01051	0.04140	0.25872
Rheumatoid arthritis (R)	GS GWEIS using ISLE - GxE effect	0.001	439	0.00388	0.04137	0.03749	-0.12605	0.04262	0.00310	0.01190	0.09394
Rheumatoid arthritis (R)	GS GWEIS using ISLE - Joint effect	0.01	3213	0.00066	0.03814	0.03749	0.05165	0.04218	0.22076	0.56634	0.82284
Rheumatoid arthritis (R)	GS GWEIS using TSLE - GxE effect	0.2	35561	0.00088	0.03836	0.03749	-0.06078	0.04348	0.16212	0.42096	0.75963
Rheumatoid arthritis (R)	GS GWEIS using TSLE - Joint effect	0.001	473	0.00424	0.04173	0.03749	0.13073	0.04210	0.00190	0.00850	0.07727
Rheumatoid arthritis (R)	UKB GWAS	0.4	56079	0.00372	0.04120	0.03749	0.12261	0.04221	0.00368	0.01940	0.14549
Rheumatoid arthritis (R)	UKB GxE effect	0.2	33861	0.00195	0.03944	0.03749	0.09018	0.04284	0.03529	0.16938	0.49819
Rheumatoid arthritis (R)	UKB GxE Joint effect	1	91009	0.00412	0.04160	0.03749	0.12975	0.04245	0.00224	0.01580	0.12153
Schizotypal personality	**GS GWAS**	**0.5**	**66399**	**0.02276**	**0.05735**	**0.03459**	**0.58398**	**0.07732**	**0.00000**	**0.00010**	**0.00200**
Schizotypal personality	GS GWEIS using DSLE - GxE effect	0.001	523	0.00083	0.03542	0.03459	0.11131	0.07822	0.15487	0.36246	0.72324
Schizotypal personality	**GS GWEIS using DSLE - Joint effect**	**0.1**	**21776**	**0.01619**	**0.05078**	**0.03459**	**0.49345**	**0.07775**	**0.00000**	**0.00010**	**0.00200**
Schizotypal personality	GS GWEIS using ISLE - GxE effect	0.01	2982	0.00279	0.03738	0.03459	0.20508	0.07837	0.00893	0.03160	0.21065
Schizotypal personality	**GS GWEIS using ISLE - Joint effect**	**0.5**	**65965**	**0.02152**	**0.05611**	**0.03459**	**0.56781**	**0.07737**	**0.00000**	**0.00010**	**0.00200**
Schizotypal	GS GWEIS using	0.005	1834	0.00108	0.03567	0.03459	0.12742	0.07824	0.10354	0.26417	0.62899

personality	TSLE - GxE effect										
Schizotypal personality	**GS GWEIS using TSLE - Joint effect**	**1**	**95511**	**0.01830**	**0.05289**	**0.03459**	**0.52391**	**0.07754**	**0.00000**	**0.00010**	**0.00200**
Schizotypal personality	**UKB GWAS**	**0.3**	**46408**	**0.00800**	**0.04259**	**0.03459**	**0.34867**	**0.07846**	**0.00001**	**0.00020**	**0.00333**
Schizotypal personality	UKB GxE effect	0.1	19908	0.00258	0.03717	0.03459	0.19968	0.07937	0.01194	0.07049	0.34997
Schizotypal personality	**UKB GxE Joint effect**	**0.4**	**55262**	**0.00915**	**0.04374**	**0.03459**	**0.37512**	**0.07890**	**0.00000**	**0.00010**	**0.00200**
Stroke	GS GWAS	0.02	5283	0.00123	0.13099	0.12976	0.11224	0.11921	0.34642	0.72313	0.88618
Stroke	GS GWEIS using DSLE - GxE effect	0.2	35891	0.00225	0.13201	0.12976	0.14540	0.10585	0.16954	0.42286	0.75963
Stroke	GS GWEIS using DSLE - Joint effect	0.02	6098	0.00205	0.13181	0.12976	0.14446	0.11643	0.21471	0.55315	0.82248
Stroke	GS GWEIS using ISLE - GxE effect	0.001	440	0.00268	0.13244	0.12976	-0.15737	0.11543	0.17279	0.44016	0.77341
Stroke	GS GWEIS using ISLE - Joint effect	0.001	421	0.00220	0.13196	0.12976	-0.14525	0.11697	0.21431	0.55574	0.82248
Stroke	GS GWEIS using TSLE - GxE effect	0.001	530	0.00884	0.13860	0.12976	-0.29457	0.11776	0.01237	0.06109	0.34997
Stroke	GS GWEIS using TSLE - Joint effect	0.005	1844	0.00177	0.13153	0.12976	-0.13614	0.12270	0.26720	0.63704	0.86475
Stroke	UKB GWAS	0.03	8340	0.00325	0.13301	0.12976	0.17140	0.11356	0.13122	0.44506	0.77341
Stroke	UKB GxE effect	0.03	7261	0.00514	0.13490	0.12976	0.22327	0.11741	0.05722	0.25917	0.62394
Stroke	UKB GxE Joint effect	0.03	8269	0.00491	0.13467	0.12976	0.21603	0.11671	0.06417	0.27347	0.63589
Stroke (R)	GS GWAS	0.04	9523	0.00096	0.01576	0.01480	0.05490	0.03510	0.11782	0.32437	0.66959
Stroke (R)	GS GWEIS using DSLE - GxE effect	0.001	523	0.00008	0.01488	0.01480	-0.01596	0.03493	0.64775	0.97800	0.99519
Stroke (R)	GS GWEIS using DSLE - Joint effect	0.001	503	0.00197	0.01677	0.01480	-0.07872	0.03508	0.02481	0.08849	0.36871
Stroke (R)	GS GWEIS using ISLE - GxE effect	0.001	441	0.00222	0.01702	0.01480	-0.08363	0.03519	0.01748	0.05889	0.34644
Stroke (R)	GS GWEIS using ISLE - Joint effect	1	95520	0.00034	0.01514	0.01480	0.03241	0.03499	0.35436	0.76572	0.89384
Stroke (R)	GS GWEIS using TSLE - GxE effect	0.001	529	0.00025	0.01505	0.01480	0.02780	0.03497	0.42659	0.82122	0.91586
Stroke (R)	GS GWEIS using	0.2	36063	0.00044	0.01524	0.01480	0.03725	0.03509	0.28851	0.65763	0.87456

	TSLE - Joint effect										
Stroke (R)	UKB GWAS	0.04	10426	0.00137	0.01617	0.01480	-0.06593	0.03519	0.06101	0.23878	0.59201
Stroke (R)	UKB GxE effect	0.001	328	0.00111	0.01591	0.01480	-0.05902	0.03508	0.09249	0.37406	0.72399
Stroke (R)	UKB GxE Joint effect	0.05	12387	0.00122	0.01602	0.01480	-0.06238	0.03539	0.07792	0.32587	0.66959

Supplementary Table 17. Genes significantly associated with depression based in gene-based analyses. Bonferroni-corrected significance threshold $p = 2.77 \times 10^{-6}$

Gene	Cohort	Effect	*p* value	Details
DCC	UK Biobank	Main additive, GWAS of PHQ	7.53×10^{-8}	*DCC* is a gene encoding for the netrin 1 receptor and is involved in neuron projection and development of prefrontal cortex during adolescence. *DCC* mediates axon guidance of neuronal growth cones towards sources of its ligand, netrin 1. We have previously shown association between the Netrin-pathway and MDD[1]. *DCC* was significantly associated with a GWAS of mood instability in UKB[2] and it is essential for NMDAR-dependent LTP and certain forms of synaptic plasticity[3]. *DCC* expression appears altered in individuals with psychiatric disorders and overexpressed in individuals who committed suicide[4]. *DCC* has been associated in a meta-analysis of depression[5], and in gene-based analyses from the two latest meta-GWAS of major depression[6,7].
ACSS3	UK Biobank	Main additive, GWAS of PHQ	6.51×10^{-7}	*ACSS3* gene encodes for an acyl-coenzyme A synthetase involved in energy generation via lipid synthesi as part of fatty acid metabolism. *ACSS3* has also been associated with increased risk of an unsustainable response to antidepressants in depressed patients[8].
	UK Biobank	Joint effect, GWEIS	1.61×10^{-6}	
DRD2	UK Biobank	Main additive, GWAS of PHQ	6.55×10^{-7}	*DRD2* encodes for the dopamine receptor D2 involved in learning and memory and has been associated with Schizophrenia[9] and cocaine dependence[10]. *DRD2* has been also reported as a significant MDD associated gene in MAGMA gene-based analyses from the two latest meta-GWAS of major depression[6,7]. *DRD2* is a major target for the action of antipsychotic drugs.
STAG1	UK Biobank	Main additive, GWAS of PHQ	1.63×10^{-6}	*STAG1* encodes for a cohesin subunit. STAG1 has been linked to schizophrenia[11] and a rare intellectual disability syndrome caused by cohesion disruption.

FOXP2	UK Biobank	Main additive, GWAS of PHQ	$2.09x10^{-6}$	*FOXP2* is expressed in the brain, regulates gene expression and is required for the correct development of language and speech[12]. *FOXP2* has been associated with educational attainment and Attention-Deficit/Hyperactivity Disorder (ADHD)[13].
KYNU	UK Biobank	Main additive, GWAS of PHQ	$2.24x10^{-6}$	*KYNU* encodes for a kynureninase. *KYNU* expression is regulated by corticosteroids via the glucocorticoid receptor and kynurenine pathway has been suggested to be involved in stress/inflammation-induced depression[14]. *KYNU* was genome-wide significant at gene-based analysis from the most recent meta-analysis of major depression[7].
PHF2	UK Biobank	Joint effect, GWEIS	$2.28x10^{-6}$	*PHF2* encodes for a histone demethylase (a Lysine demethylase that also demethylates non-histone proteins) involved in epigenetic regulation of gene expression. *PHF2* has been identified as tumour suppressor in some cancers[15]. *PHF2* was genome-wide significant at gene-based analysis from the most recent meta-analysis of major depression[7].
MTNR1B	Generation Scotland	GxE, GWEIS using DSLE	$1.53x10^{-6}$	*MTNR1B* encodes for the melatonin receptor 1B. This receptor is involved in circadian rhythms of which disruption may lead to cardiometabolic disorders and type 2 diabetes (T2D)[16]. Several polymorphisms in *MTNR1B* have been associated with blood pressure, cardiac parameters, fasting plasma glucose levels and T2D[17-20]. *MNTNR1B* is targeted by the neurohormone melatonin and melatonin agonists have been reported to produce antidepressant effects[21].
	Generation Scotland	Joint effect, GWEIS using DSLE	$2.38x10^{-6}$	

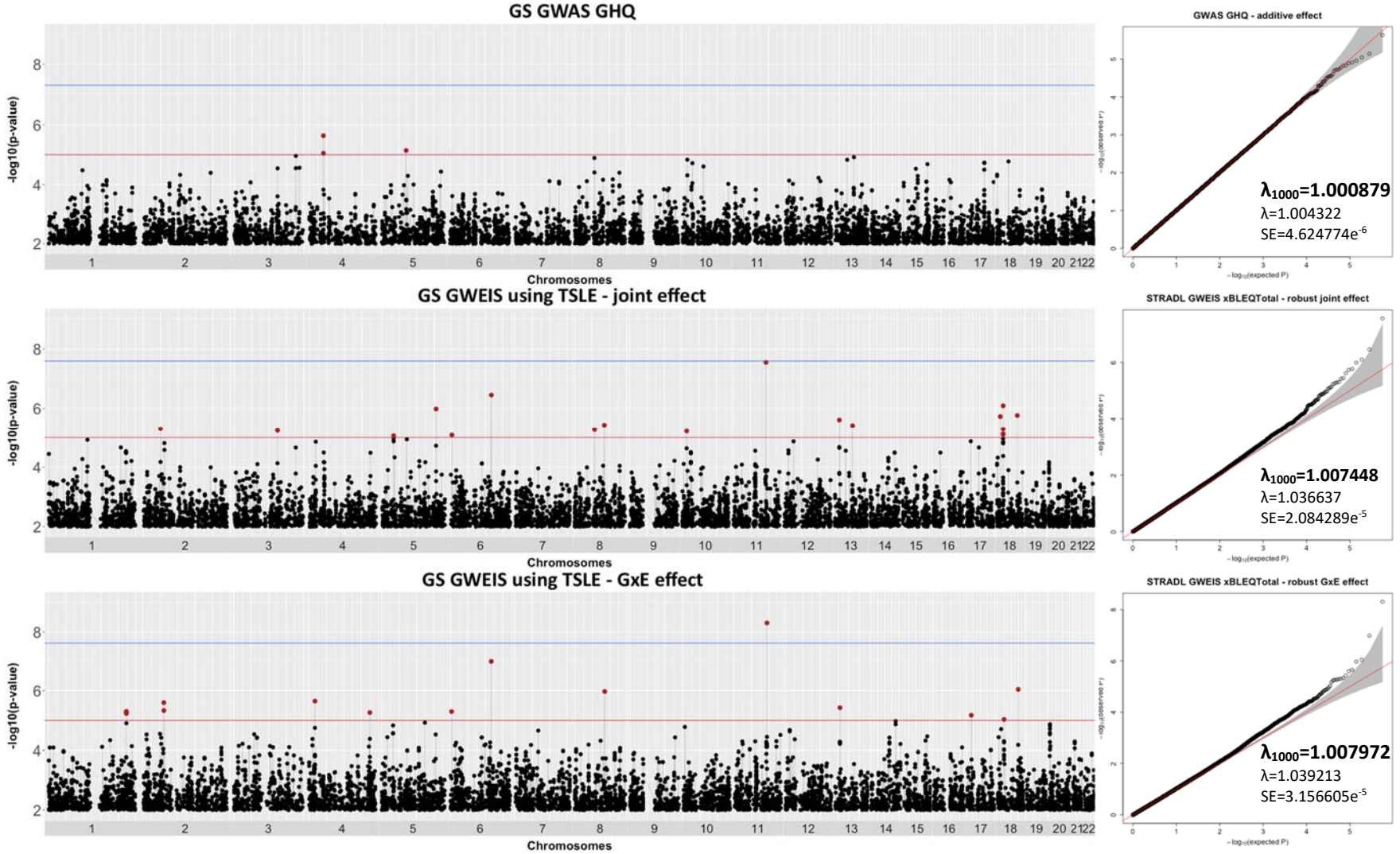
References

1. Zeng, Y. et al. A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway Is Associated With Major Depressive Disorder. Biol Psychiatry 81, 336-346 (2017).
2. Ward, J. et al. Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. Transl Psychiatry 7, 1264 (2017).
3. Horn, K.E. et al. DCC expression by neurons regulates synaptic plasticity in the adult brain. Cell Rep 3, 173-85 (2013).
4. Manitt, C. et al. dcc orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. Transl Psychiatry 3, e338 (2013).
5. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide

analyses. Nat Genet 48, 624-33 (2016).
6. Wray, N.R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50, 668-681 (2018).
7. Howard, D.M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22, 343-352 (2019).
8. Hunter, A.M. et al. A genome-wide association study of a sustained pattern of antidepressant response. J Psychiatr Res 47, 1157-65 (2013).
9. Zhang, J.P. et al. Association of a Schizophrenia Risk Variant at the DRD2 Locus With Antipsychotic Treatment Response in First-Episode Psychosis. Schizophr Bull 41, 1248-55 (2015).
10. Sullivan, D. et al. Dopamine transporter DAT and receptor DRD2 variants affect risk of lethal cocaine abuse: a gene-gene-environment interaction. Transl Psychiatry 3, e222 (2013).
11. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-7 (2014).
12. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413, 519-23 (2001).
13. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
14. Brooks, A.K. et al. Interactions between inflammatory mediators and corticosteroids regulate transcription of genes within the Kynurenine Pathway in the mouse hippocampus. J Neuroinflammation 13, 98 (2016).
15. Lee, K.H. et al. PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer. Oncogene 34, 2897-909 (2015).
16. Maury, E., Ramsey, K.M. & Bass, J. Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. Circ Res 106, 447-62 (2010).
17. Bouatia-Naji, N. et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet 41, 89-94 (2009).
18. Prokopenko, I. et al. Variants in MTNR1B influence fasting glucose levels. Nat Genet 41, 77-81 (2009).
19. Lyssenko, V. et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet 41, 82-8 (2009).
20. Huber, M. et al. Genetics of melatonin receptor type 2 is associated with left ventricular function in hypertensive patients treated according to guidelines. Eur J Intern Med 24, 650-5 (2013).
21. Dubovsky, S.L. & Warren, C. Agomelatine, a melatonin agonist with antidepressant properties. Expert Opin Investig Drugs 18, 1533-40 (2009).

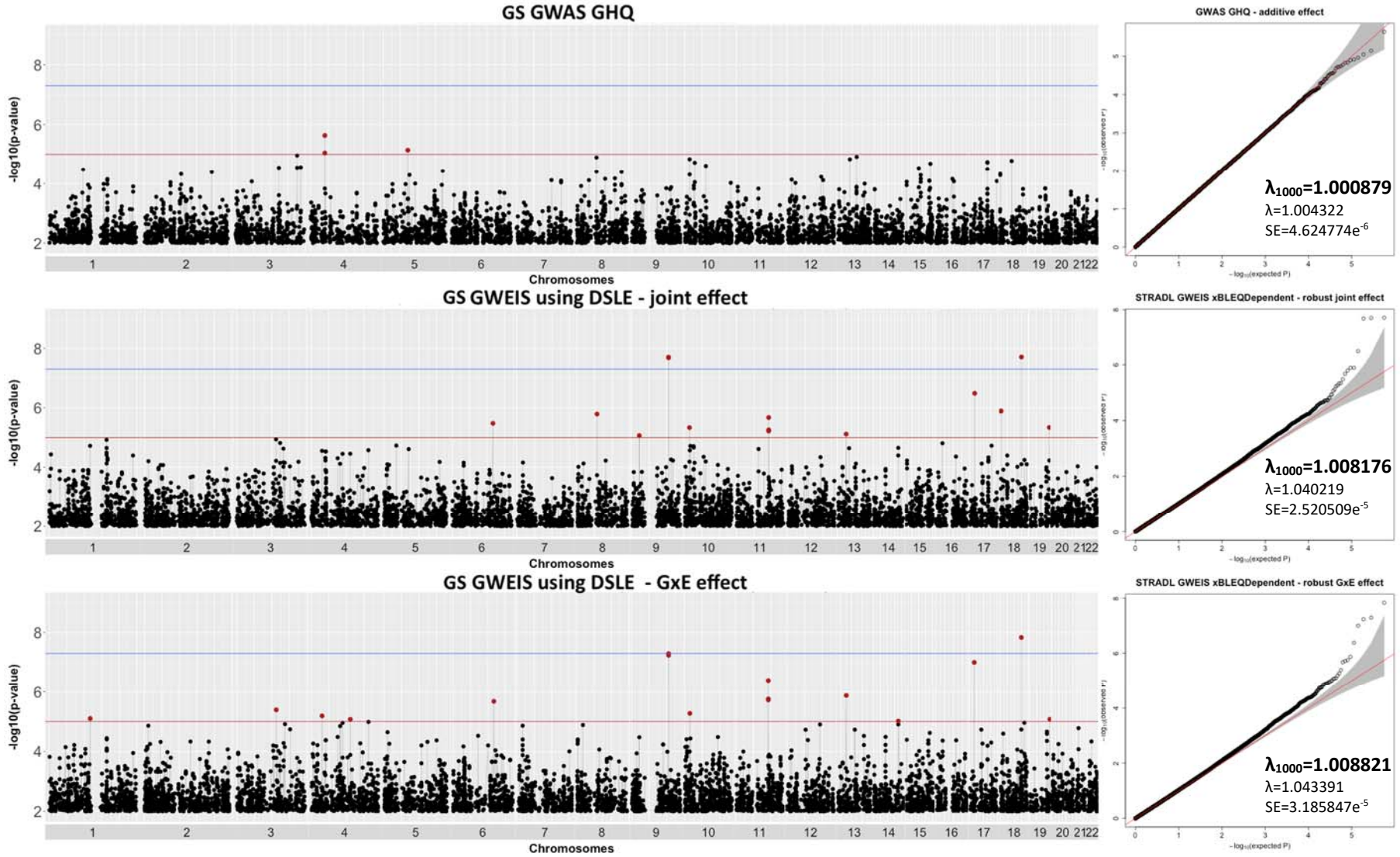
D.6 Supplementary Figures

Supplementary Figure 1. Manhattan and QQ plots of Generation Scotland GWAS and GWEIS using TSLE as exposure



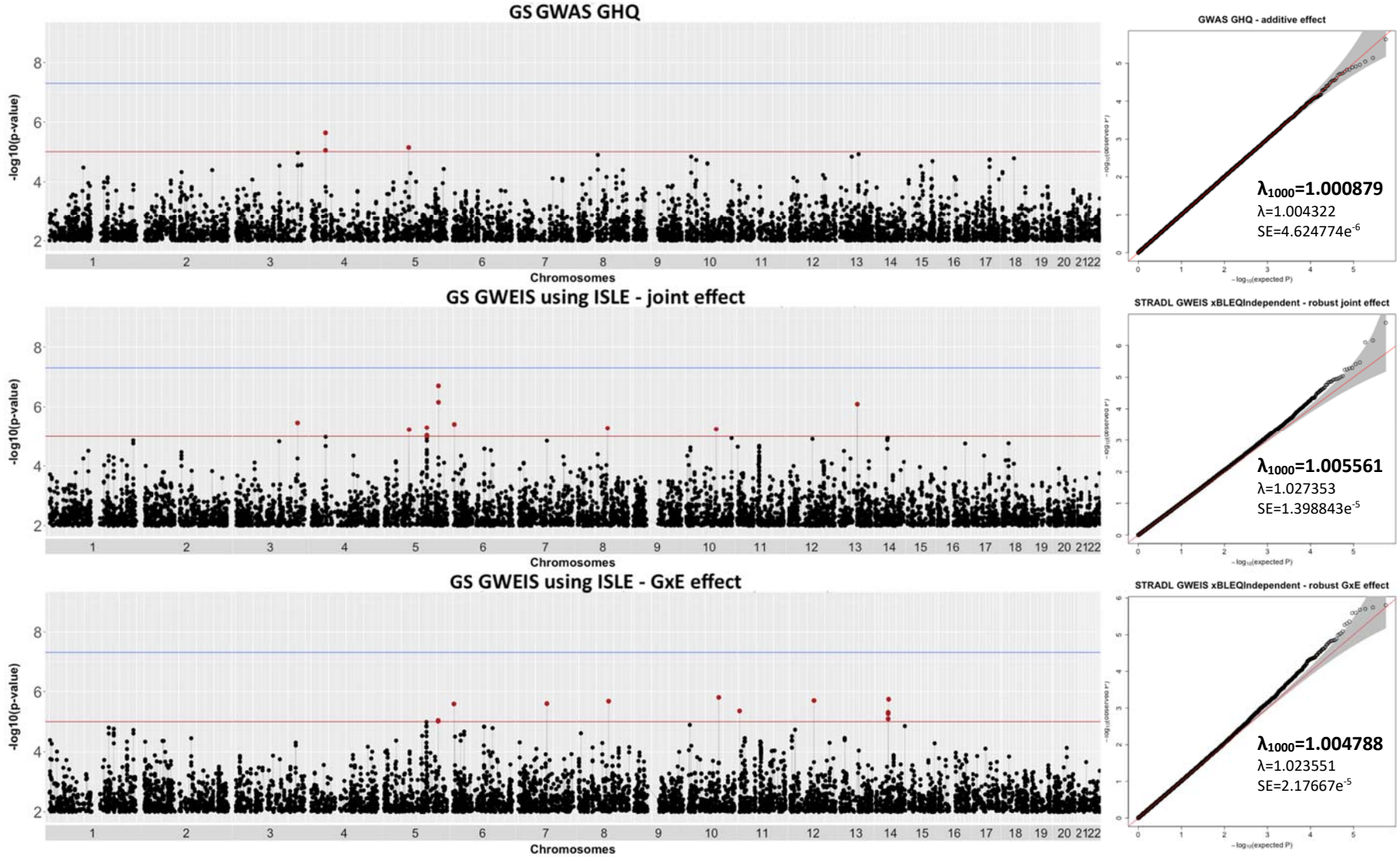
Manhattan plots showing associations of GHQ in Generation Scotland (N = 4,919) from (top) additive effects from GWAS, (middle) joint effect from GWEIS and (bottom) GxE effect from GWEIS, using "total" SLE (TSLE) as exposure. Suggestive genome-wide significance threshold (p = 1x10^{-5}) is shown by solid red line. Genome-wide significance threshold (GWAS: p = 5x10^{-8}; GWEIS: p = 2.97x10^{-8}) is shown by solid blue line. QQ plots on the right show genomic factor inflation and represented by lambda estimates λ and λ_{1000}; grey shadow represents the 95% confidence intervals.

Supplementary Figure 2. Manhattan and QQ plots of Generation Scotland GWAS and GWEIS using DSLE as exposure



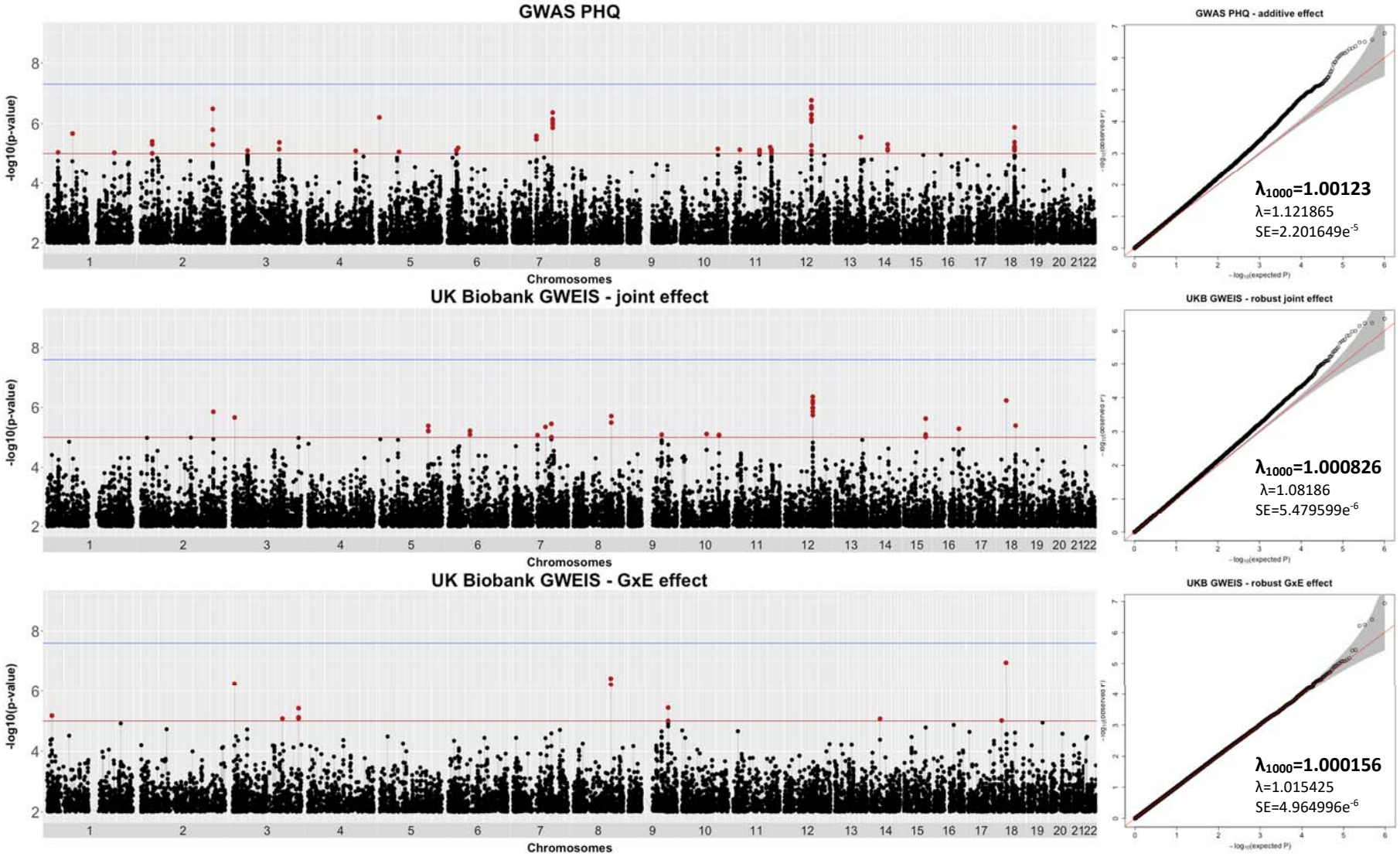
Manhattan plots showing associations of GHQ in Generation Scotland (N = 4,919) from (top) additive effects from GWAS, (middle) joint effect from GWEIS and (bottom) GxE effect from GWEIS, using "dependent" SLE (DSLE) as exposure. Suggestive genome-wide significance threshold (p = 1x10^{-5}) is shown by solid red line. Genome-wide significance threshold (GWAS: p = 5x10^{-8}; GWEIS: p = 2.97x10^{-8}) is shown by solid blue line. QQ plots on the right show genomic factor inflation and represented by lambda estimates λ and λ_{1000}; grey shadow represents the 95% confidence intervals.

Supplementary Figure 3. Manhattan and QQ plots of Generation Scotland GWAS and GWEIS using ISLE as exposure



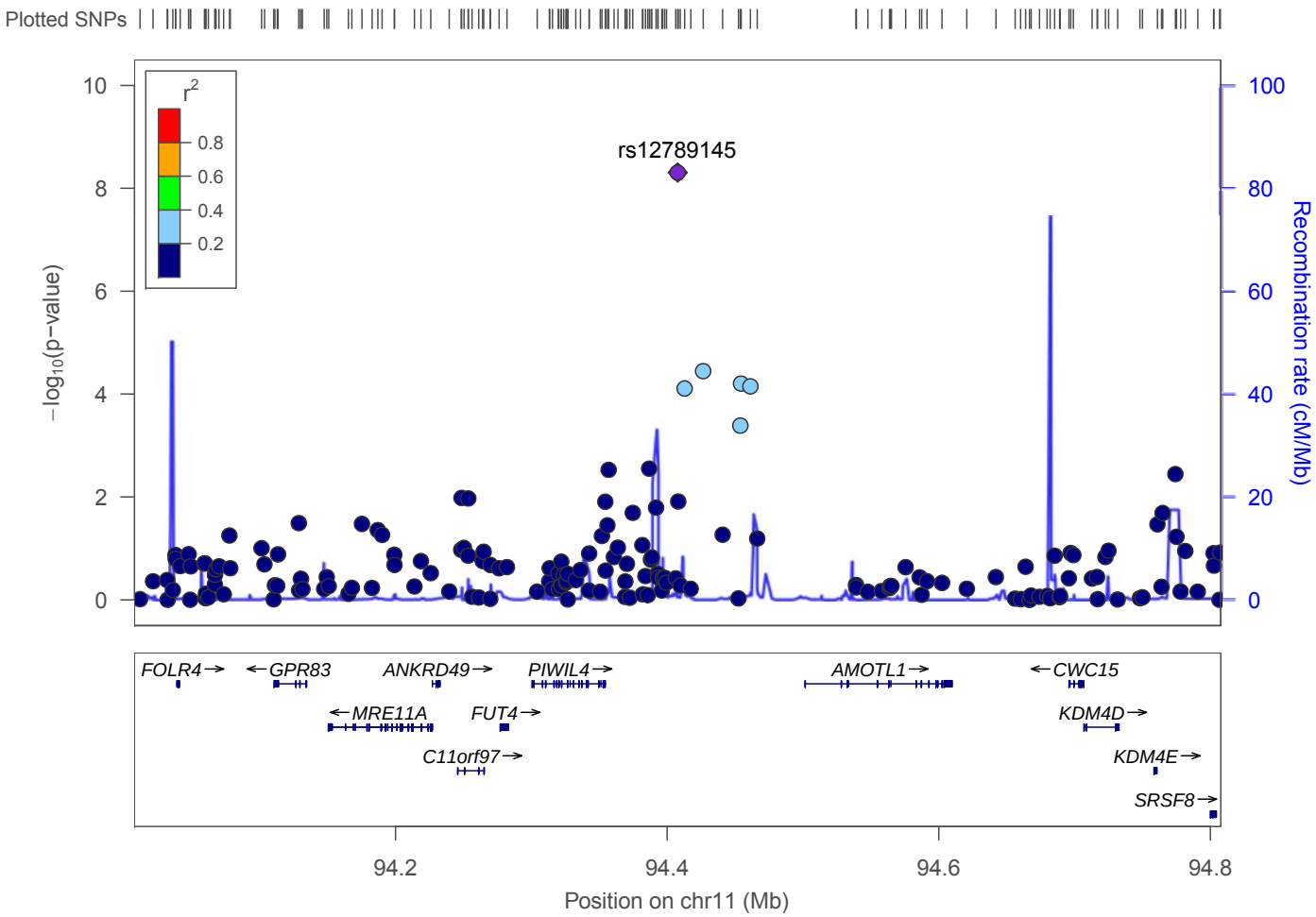
Manhattan plots showing associations of GHQ in Generation Scotland (N = 4,919) from (top) additive effects from GWAS, (middle) joint effect from GWEIS and (bottom) GxE effect from GWEIS, using "independent" SLE (ISLE) as exposure. Suggestive genome-wide significance threshold (p = 1x10^{-5}) is shown by solid red line. Genome-wide significance threshold (GWAS: p = 5x10^{-8}; GWEIS: p = 2.97x10^{-8}) is shown by solid blue line. QQ plots on the right show genomic factor inflation and represented by lambda estimates λ and λ_{1000}; grey shadow represents 95% confidence intervals.

Supplementary Figure 4. Manhattan and QQ plots of UK Biobank GWAS and GWEIS



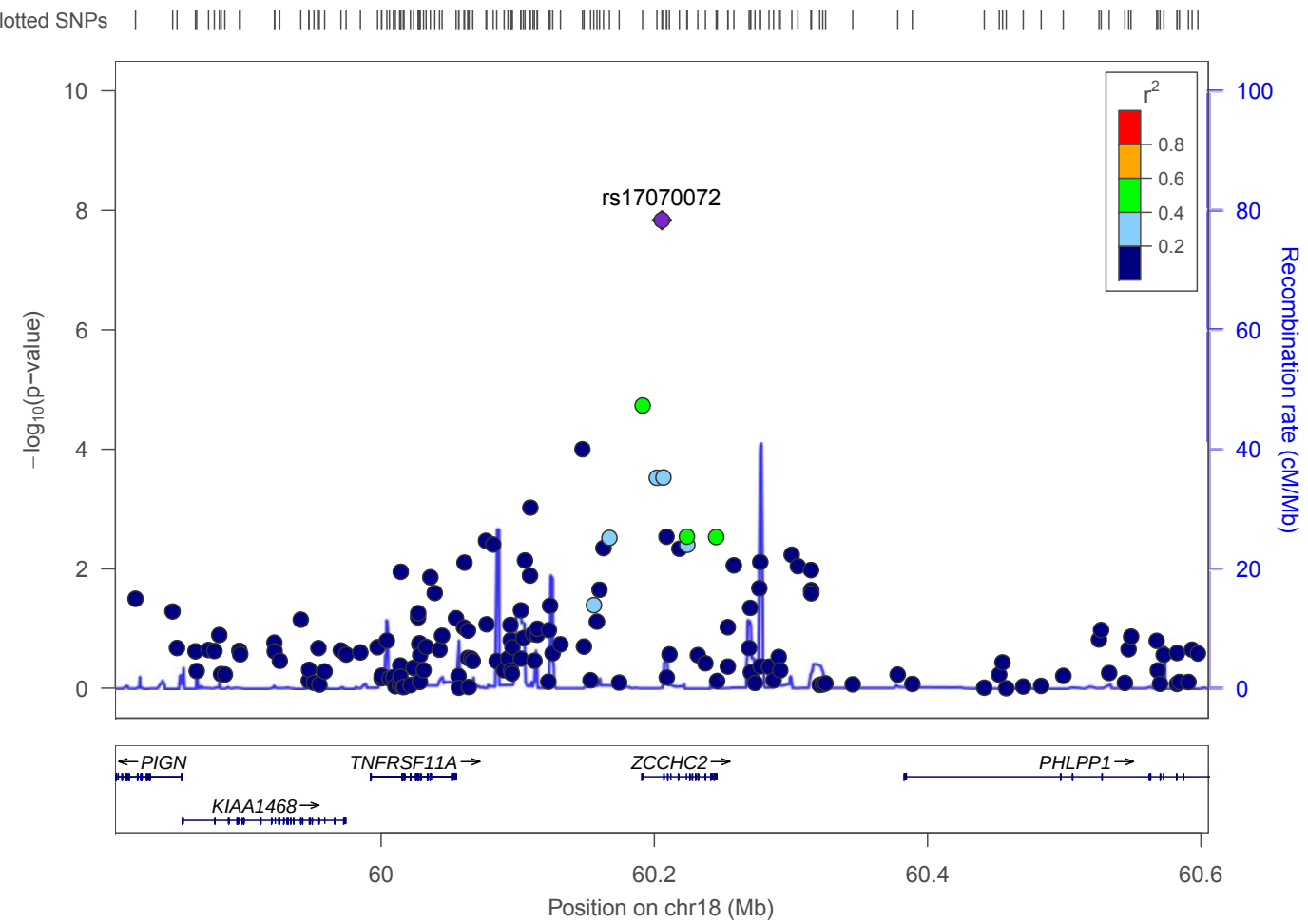
Manhattan plots showing associations of PHQ in UK Biobank (N = 99,057) from (top) additive effects from GWAS, (middle) joint effect from GWEIS and (bottom) GxE effect from GWEIS, using "total" SLE as exposure. Suggestive genome-wide significance threshold (p = 1x10^{-5}) is shown by solid red line. Genome-wide significance threshold (GWAS: p = 5x10^{-8}; GWEIS: p = 2.47x10^{-8}) is shown by solid blue line. QQ plots on the right show genomic factor inflation and represented by lambda estimates λ and λ_{1000}; grey shadow represents the 95% confidence intervals.

Supplementary Figure 5. GxE effect derived from GWEIS on Generation Scotland using TSLE as exposure, regional plot of locus around rs12789145 on chromosome 11



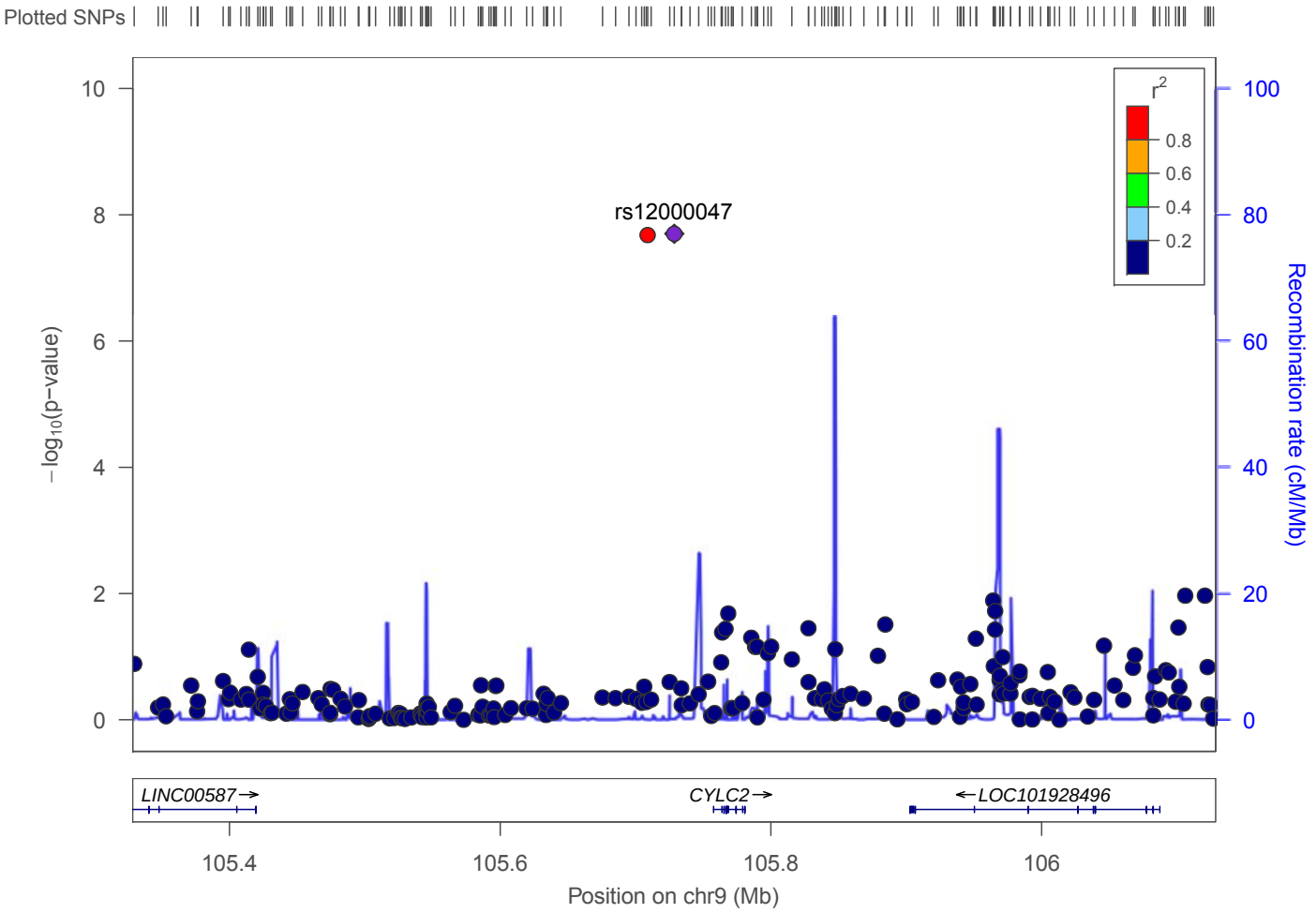
LocusZoom plot of the associated region on chromosome 11. Genotyped SNPs are shown as circles. Linkage disequilibrium structure across the region is represented by colour sclae (r^2). Recombination rate is indicated by solid blue line. Y-axis shows significance association with GxE effect derived from GWEIS on Generation Scotland using TSLE as exposure. Location, structure and transcription direction of the genes in the region are represented below.

Supplementary Figure 6. GxE effect derived from GWEIS on Generation Scotland using DSLE as exposure, regional plot of locus around rs17070072 on chromosome 18



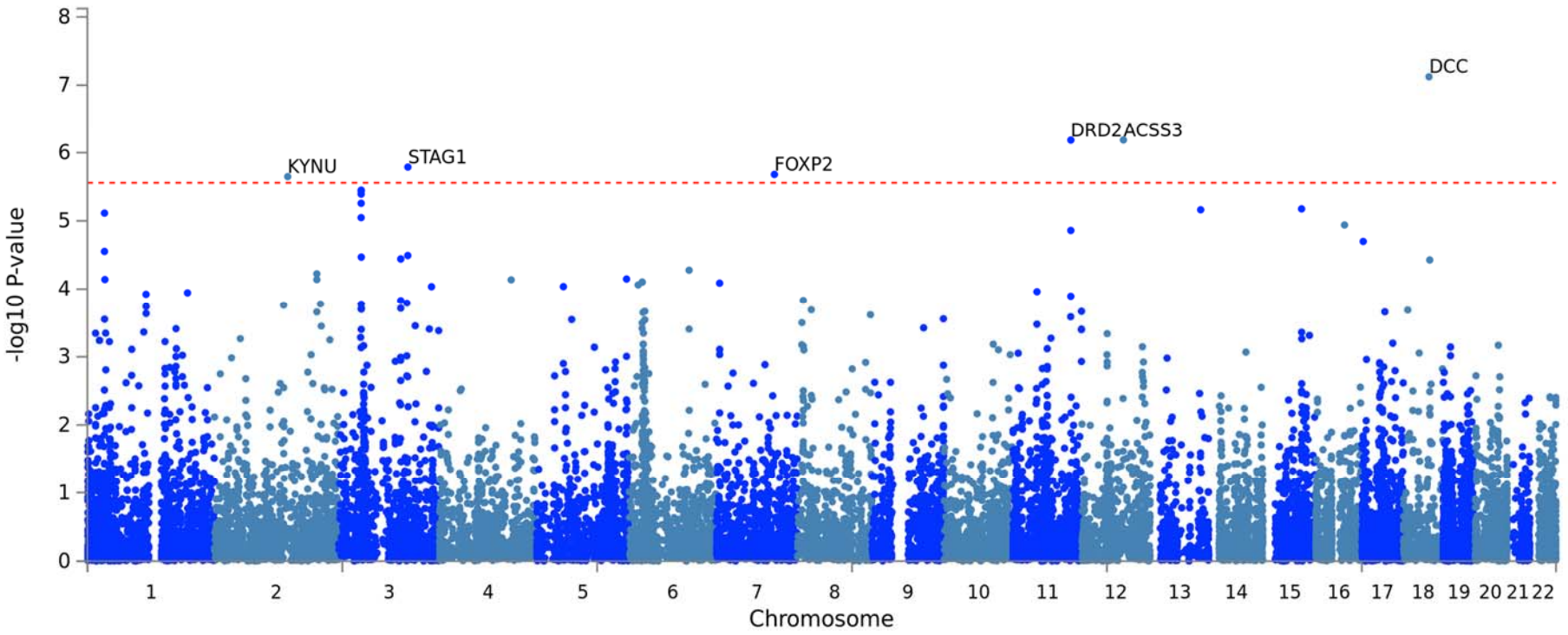
LocusZoom plot of the associated region on chromosome 18. Genotyped SNPs are shown as circles. Linkage disequilibrium structure across the region is represented by colour sclae (r^2). Recombination rate is indicated by solid blue line. Y-axis shows significance association with GxE effect derived from GWEIS on Generation Scotland using DSLE as exposure. Location, structure and transcription direction of the genes in the region are represented below.

Supplementary Figure 7. Joint effect derived from GWEIS on Generation Scotland using DSLE as exposure, regional plot of locus around rs17070072 on chromosome 9



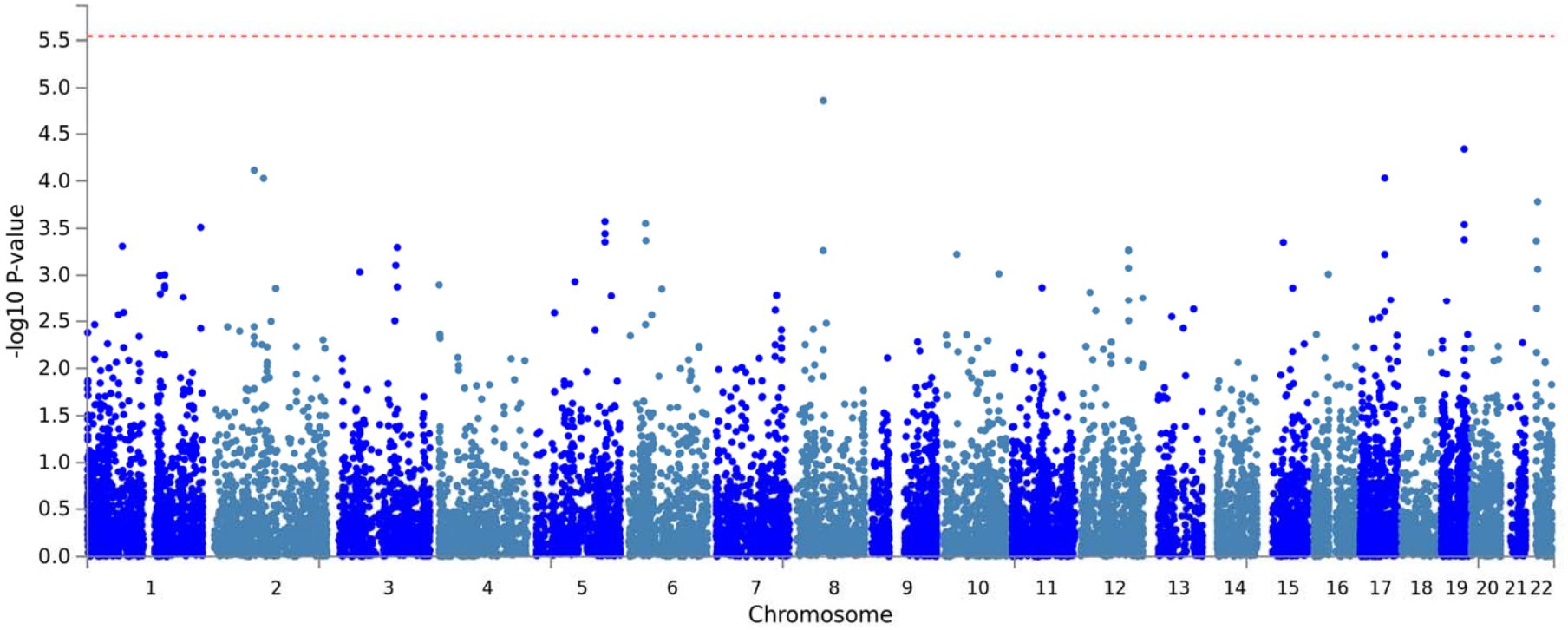
LocusZoom plot of the associated region on chromosome 9. Genotyped SNPs are shown as circles. Linkage disequilibrium structure across the region is represented by colour sclae (r^2). Recombination rate is indicated by solid blue line. Y-axis shows significance association with joint effect derived from GWEIS on Generation Scotland using DSLE as exposure. Location, structure and transcription direction of the genes in the region are represented below.

Supplementary Figure 8. Manhattan plot of gene-based test of PHQ UKB GWAS



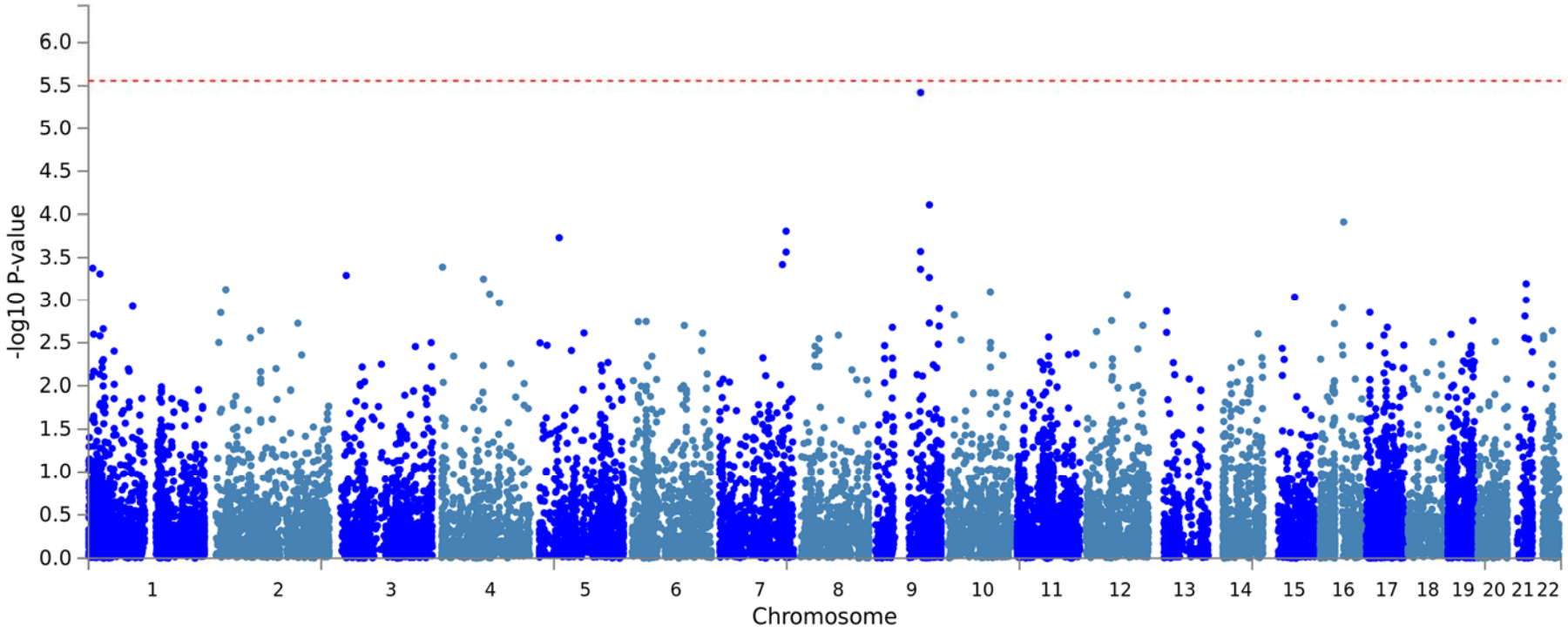
Manhattan plot showing gene-based association of PHQ using summary statistics for the additive effect derived from GWAS in UK Biobank (N = 99,057). The x-axis is base-paired chromosomal position and the y-axis represents the significance ($-\log_{10}$ *p* value) of association between main additive effects and PHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 9. Manhattan plot of gene-based test of GHQ GS GWAS



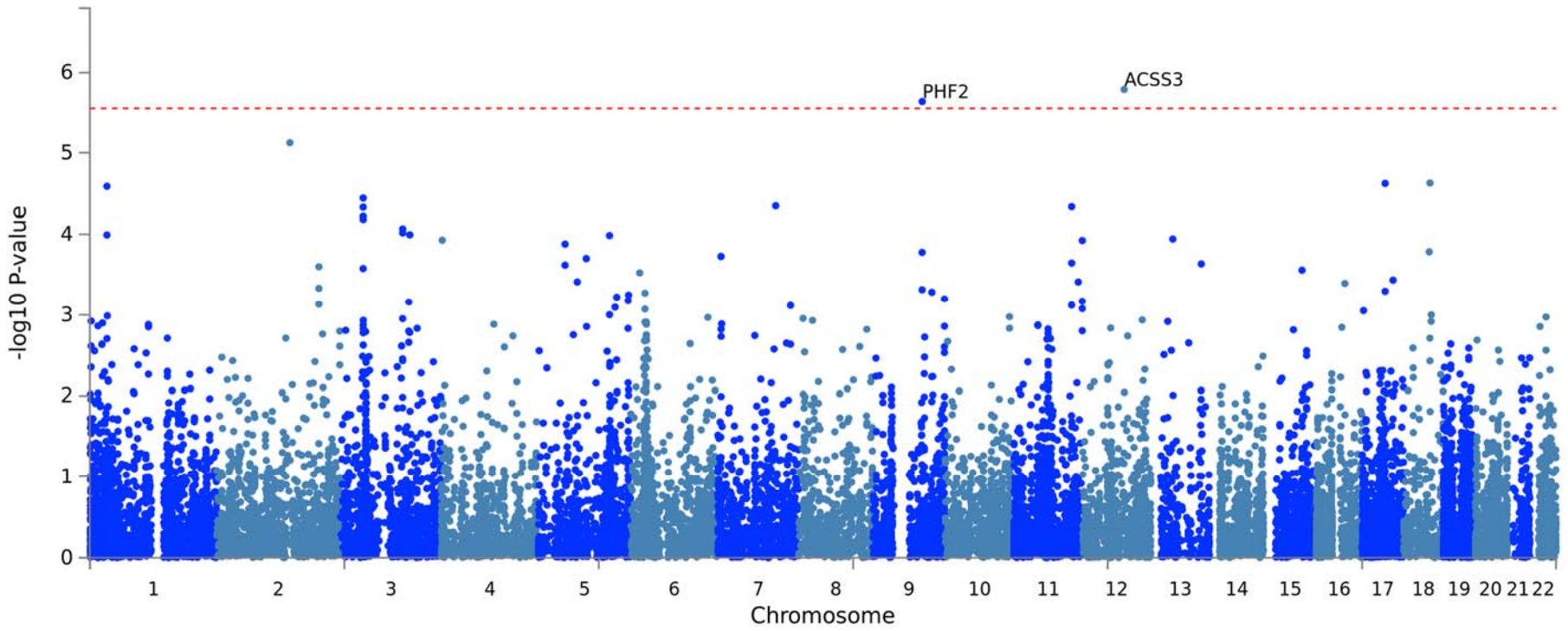
Manhattan plot showing gene-based association of GHQ using summary statistics for the additive effect derived from GWAS in Generation Scotland (N = 4,919). The x-axis is base-paired chromosomal position and the y-axis represents the significance ($-\log_{10} p$ value) of association between main additive effects and GHQ. Bonferroni-corrected significance threshold (p = 0.05 / 18,068 = 2.77×10^{-6}) is shown by dashed red line.

Supplementary Figure 10. Manhattan plot of gene-based test of PHQ UKB GWEIS using $TSLE_{UKB}$ as exposure for GxE effect



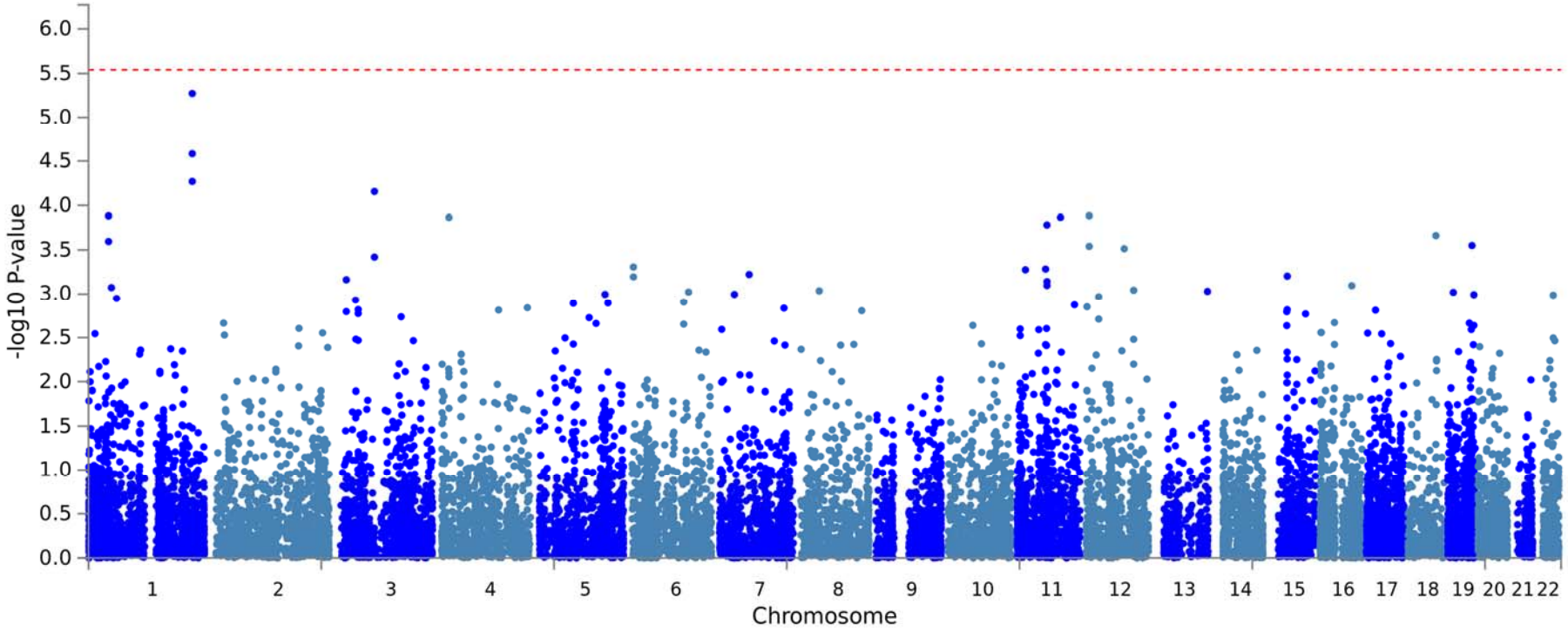
Manhattan plot showing gene-based association of PHQ using summary statistics for the GxE effect derived from GWEIS in UK Biobank (N = 99,057). The x-axis is base-paired chromosomal position and the y-axis represents the significance ($-\log_{10}$ *p* value) of association between GxE effects and PHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 11. Manhattan plot of gene-based test of PHQ UKB GWEIS using $TSLE_{UKB}$ as exposure for joint effect



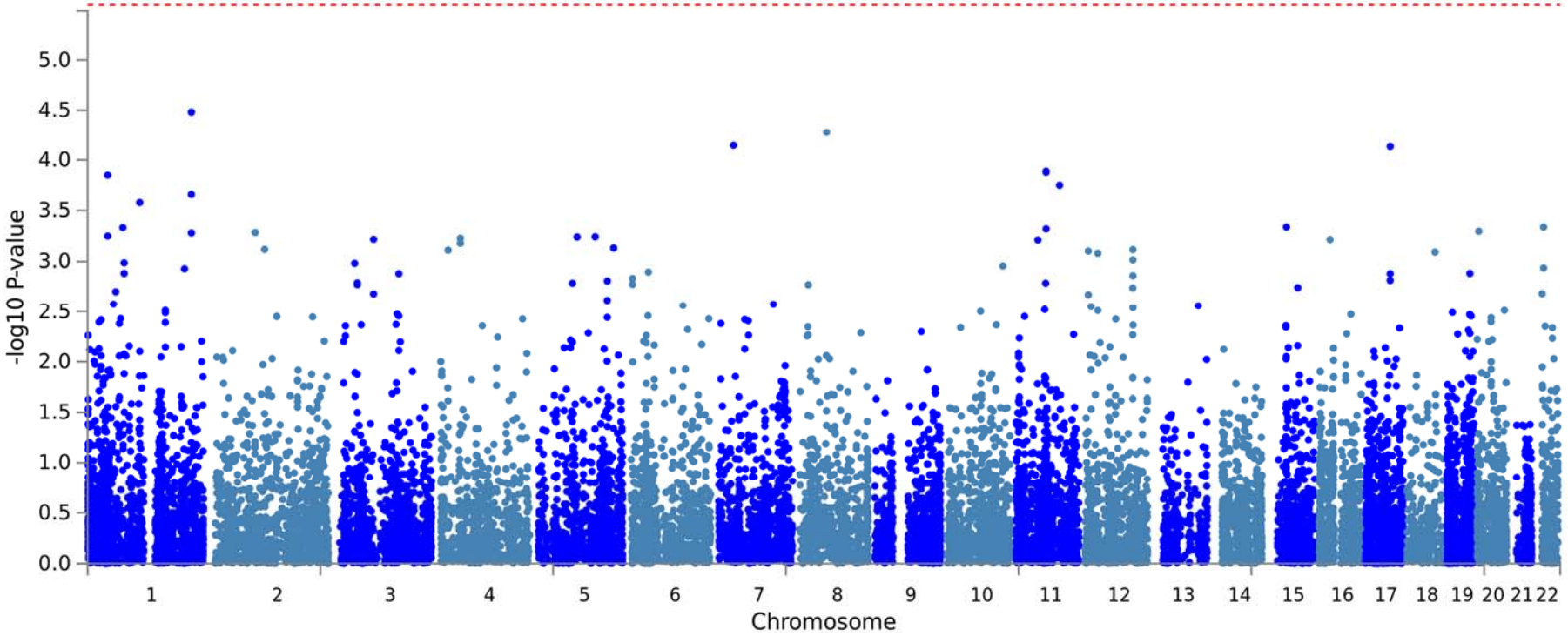
Manhattan plot showing gene-based association of PHQ using summary statistics for the joint effect derived from GWEIS in UK Biobank (N = 99,057). The x-axis is base-paired chromosomal position and the y-axis represents the significance ($-\log_{10}$ *p* value) of association between joint effects and PHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 12. Manhattan plot of gene-based test of GHQ GS GWEIS using TSLE as exposure for GxE effect



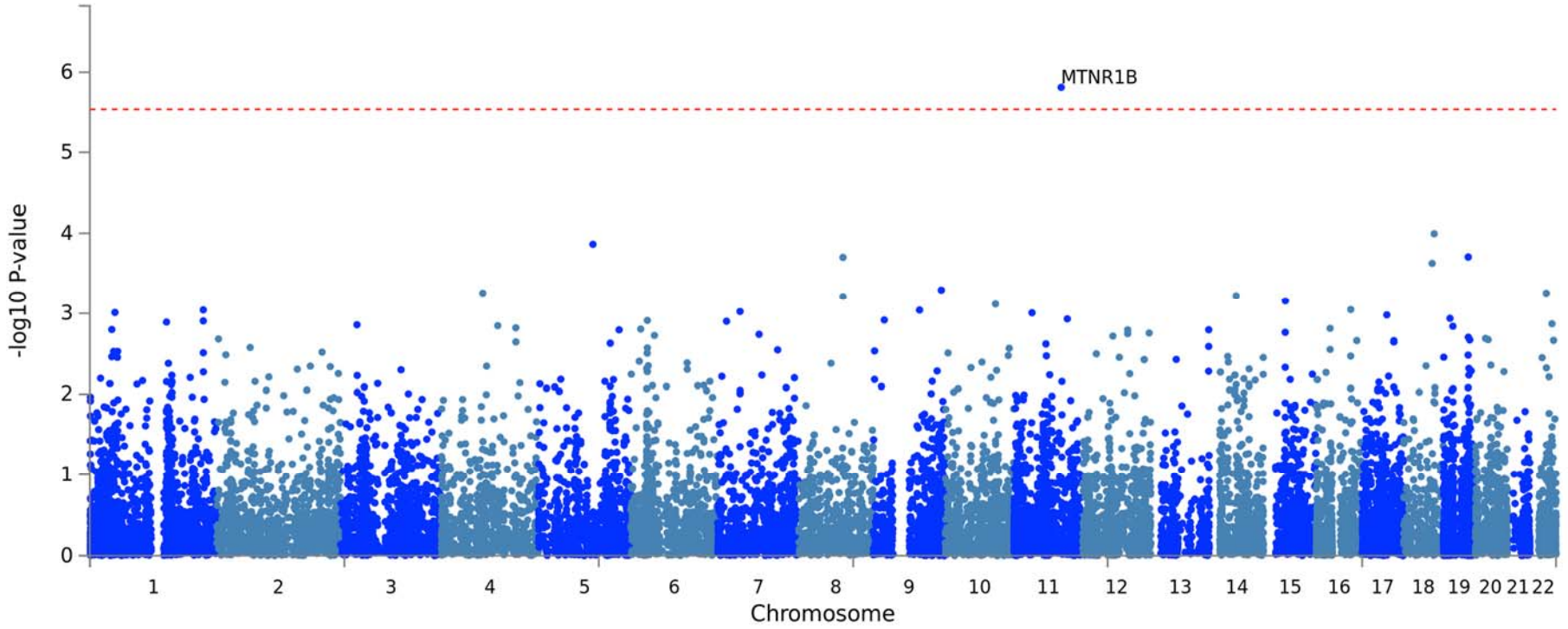
Manhattan plot showing gene-based association of GHQ using summary statistics for the GxE effect derived from GWEIS in Generation Scotland (N = 4,919) using TSLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance (-$\log_{10}$ *p* value) of association between GxE effects and GHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 13. Manhattan plot of gene-based test of GHQ GS GWEIS using TSLE as exposure for joint effect



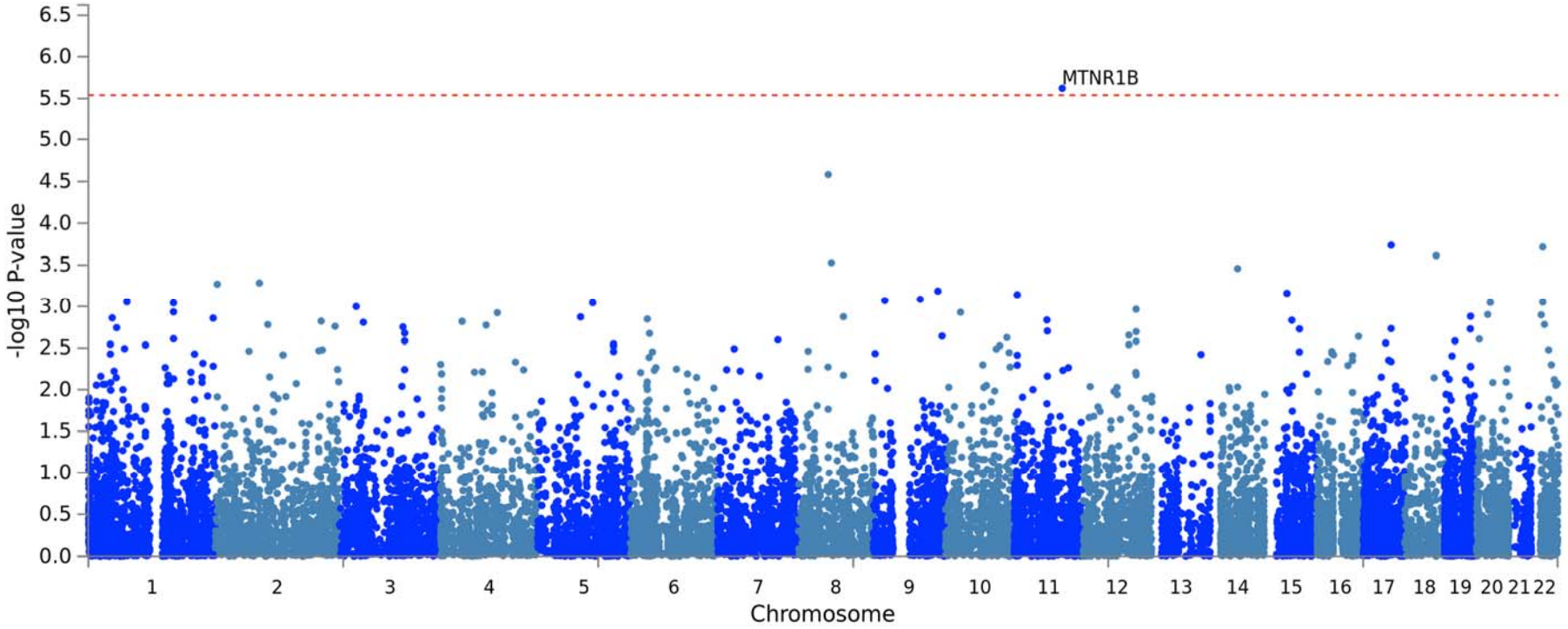
Manhattan plot showing gene-based association of GHQ using summary statistics for the joint effect derived from GWEIS in Generation Scotland (N = 4,919) using TSLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance (-$\log_{10}$ *p* value) of association between joint effects and GHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 14. Manhattan plot of gene-based test of GHQ GS GWEIS using DSLE as exposure for GxE effect



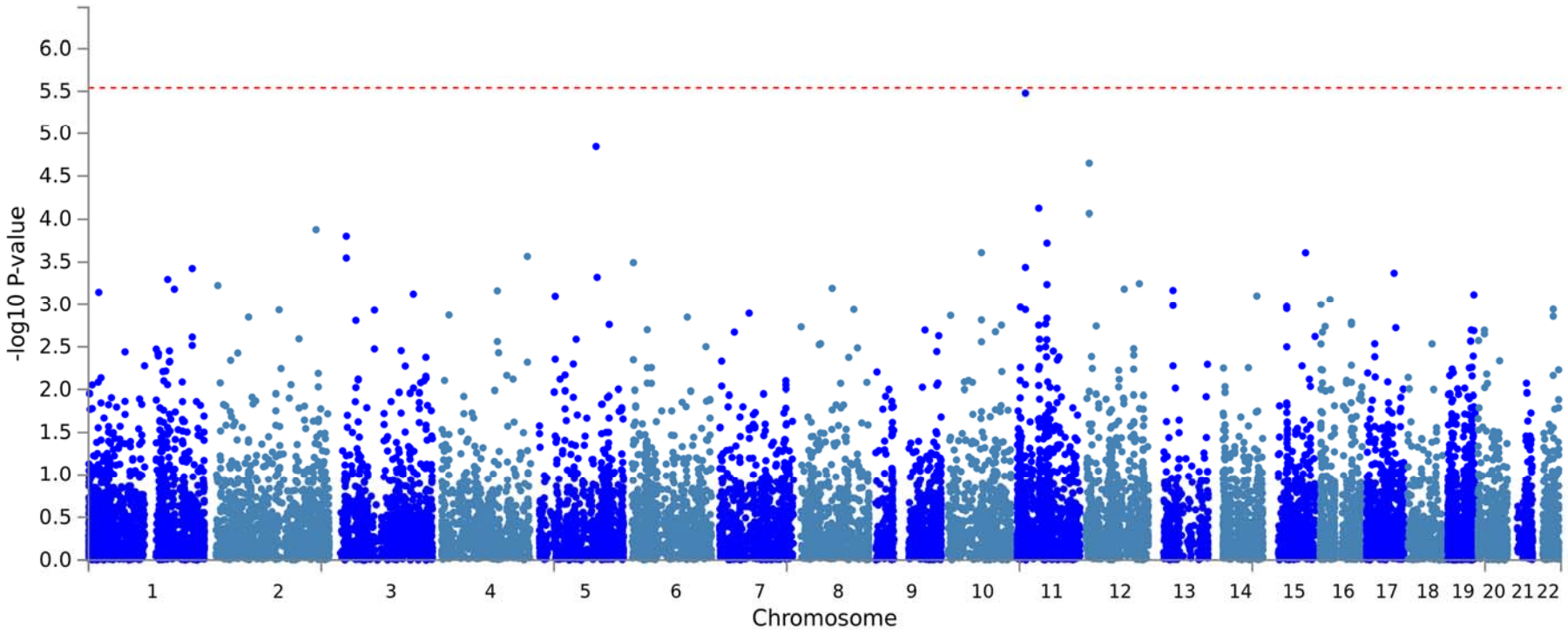
Manhattan plot showing gene-based association of GHQ using summary statistics for the GxE effect derived from GWEIS in Generation Scotland (N = 4,919) using DSLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance (-$\log_{10}$ *p* value) of association between GxE effects and GHQ. Bonferroni-corrected significance threshold (*p* = 0.05 / 18,068 = 2.77×10^{-6}) is shown by dashed red line.

Supplementary Figure 15. Manhattan plot of gene-based test of GHQ GS GWEIS using DSLE as exposure for joint effect



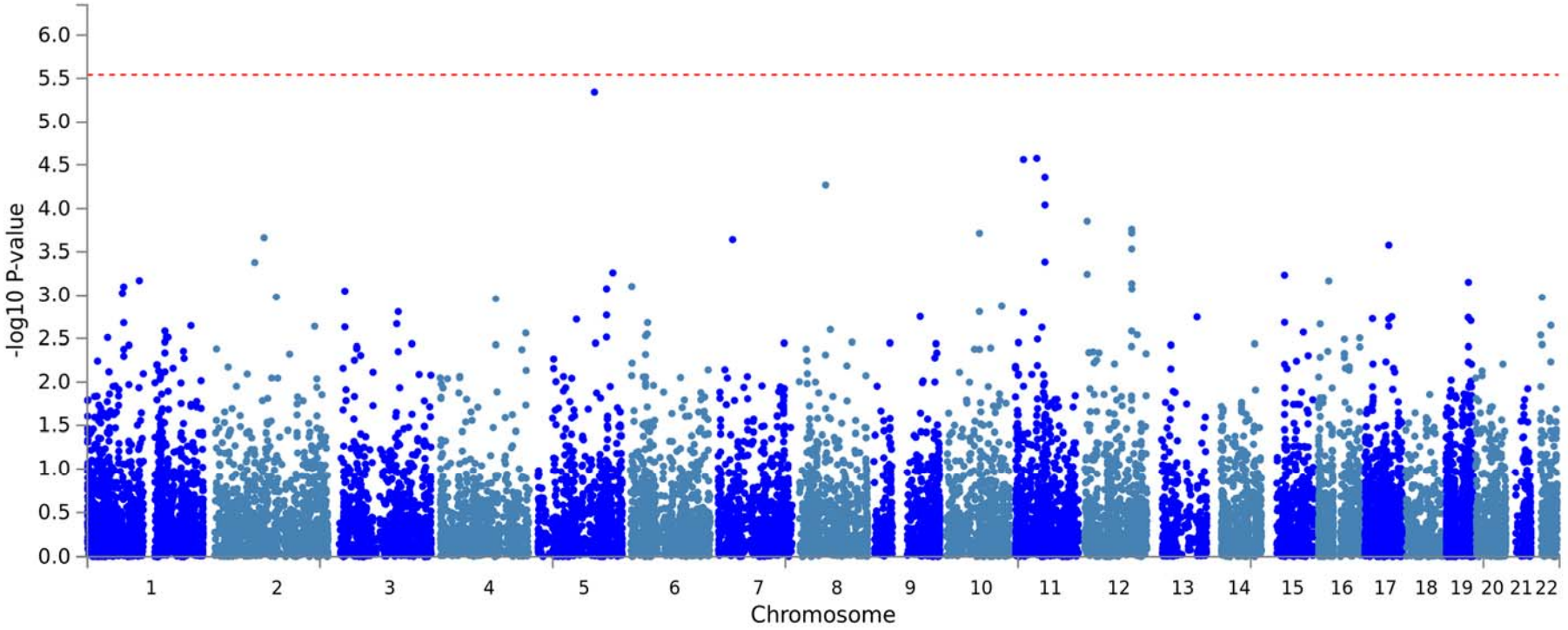
Manhattan plot showing gene-based association of GHQ using summary statistics for the joint effect derived from GWEIS in Generation Scotland using DSLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance ($-\log_{10}$ *p* value) of association between joint effects and GHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 16. Manhattan plot of gene-based test of GHQ GS GWEIS using ISLE as exposure for GxE effect



Manhattan plot showing gene-based association of GHQ using summary statistics for the joint effect derived from GWEIS in Generation Scotland (N = 4,919) using ISLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance (-$\log_{10}$ *p* value) of association between joint effects and GHQ. Bonferroni-corrected significance threshold ($p = 0.05 / 18{,}068 = 2.77 \times 10^{-6}$) is shown by dashed red line.

Supplementary Figure 17. Manhattan plot of gene-based test of GHQ GS GWEIS using ISLE as exposure for joint effect



Manhattan plot showing gene-based association of GHQ using summary statistics for the joint effect derived from GWEIS in Generation Scotland (N = 4,919) using ISLE as exposure. The x-axis is base-paired chromosomal position and the y-axis represents the significance (-$\log_{10}$ *p* value) of association between joint effects and GHQ. Bonferroni-corrected significance threshold (*p* = 0.05 / 18,068 = 2.77×10^{-6}) is shown by dashed red line.

D.7 Arnau-Soler *et al.*, 2019, Translational Psychiatry (II)

Arnau-Soler et al. *Translational Psychiatry* (2019)9:14
https://doi.org/10.1038/s41398-018-0360-y

Translational Psychiatry

ARTICLE Open Access

Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland

Aleix Arnau-Soler[1], Erin Macdonald-Dunlop[2], Mark J. Adams[3], Toni-Kim Clarke[3], Donald J. MacIntyre[3], Keith Milburn[4], Lauren Navrady[3], Generation Scotland[5], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Caroline Hayward, Andrew M. McIntosh[3] and Pippa A. Thomson[1]

Abstract

Stress is associated with poorer physical and mental health. To improve our understanding of this link, we performed genome-wide association studies (GWAS) of depressive symptoms and genome-wide by environment interaction studies (GWEIS) of depressive symptoms and stressful life events (SLE) in two UK population-based cohorts (Generation Scotland and UK Biobank). No SNP was individually significant in either GWAS, but gene-based tests identified six genes associated with depressive symptoms in UK Biobank (*DCC, ACSS3, DRD2, STAG1, FOXP2 and KYNU;* $p < 2.77 \times 10^{-6}$). Two SNPs with genome-wide significant GxE effects were identified by GWEIS in Generation Scotland: rs12789145 (53-kb downstream *PIWIL4*; $p = 4.95 \times 10^{-9}$; total SLE) and rs17070072 (intronic to *ZCCHC2*; $p = 1.46 \times 10^{-8}$; dependent SLE). A third locus upstream *CYLC2* (rs12000047 and rs12005200, $p < 2.00 \times 10^{-8}$; dependent SLE) when the joint effect of the SNP main and GxE effects was considered. GWEIS gene-based tests identified: *MTNR1B* with GxE effect with dependent SLE in Generation Scotland; and *PHF2* with the joint effect in UK Biobank ($p < 2.77 \times 10^{-6}$). Polygenic risk scores (PRSs) analyses incorporating GxE effects improved the prediction of depressive symptom scores, when using weights derived from either the UK Biobank GWAS of depressive symptoms ($p = 0.01$) or the PGC GWAS of major depressive disorder ($p = 5.91 \times 10^{-3}$). Using an independent sample, PRS derived using GWEIS GxE effects provided evidence of shared aetiologies between depressive symptoms and schizotypal personality, heart disease and COPD. Further such studies are required and may result in improved treatments for depression and other stress-related conditions.

Introduction

Mental illness results from the interplay between genetic susceptibility and environmental risk factors[1,2]. Previous studies have shown that the effects of environmental factors on traits may be partially heritable[3] and moderated by genetics[4,5]. Major depressive disorder (MDD) is the most common psychiatric disorder with a lifetime prevalence of approximately 14% globally[6] and with a heritability of approximately 37%[7]. There is strong evidence for the role of stressful life events (SLEs) as risk factor and trigger for

Correspondence: Aleix Arnau-Soler (aleix.arnau.soler@igmm.ed.ac.uk) or Pippa A. Thomson (Pippa.Thomson@ed.ac.uk)
[1]Medical Genetics Section, University of Edinburgh, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK
[2]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK
Full list of author information is available at the end of the article.
Generation Scotland is a collaboration between the University Medical School and NHS in Aberdeen, Dundee, Edinburgh and Glasgow, Scotland, UK

depression[8–12]. Genetic control of sensitivity to stress may vary between individuals, resulting in individual differences in the depressogenic effects of SLE, i.e., genotype-by-environment interaction (GxE)[4,13–16]. Significant evidence of GxE has been reported for common respiratory diseases and some forms of cancer[17–22], and GxE studies have identified genetic risk variants not found by genome-wide association studies (GWAS)[23–27].

Interaction between polygenic risk of MDD and recent SLE are reported to increase liability to depressive symptoms[4,16]; validating the implementation of genome-wide approaches to study GxE in depression. Most GxE studies for MDD have been conducted on candidate genes, or using polygenic approaches to a wide range of environmental risk factors, with some contradictory findings[28–32]. Incorporating knowledge about recent SLE into GWAS may improve our ability to detect risk variants in depression otherwise missed in GWAS[33]. To date, three studies have performed genome-wide by environment interaction studies (GWEIS) of MDD and SLE[34–36], but this is the first study to perform GWEIS of depressive symptoms using adult SLE in cohorts of relatively homogeneous European ancestry.

Interpretation of GxE effects may be hindered by gene–environment correlation. Gene–environment correlation denotes a genetic mediation of associations through genetic influences on exposure to, or reporting of, environments[2,37]. Genetic factors predisposing to MDD may contribute to exposure and/or reporting of SLE[38]. To tackle this limitation, measures of SLE can be broken down into SLE likely to be independent of a respondent's own behaviour and symptoms, or into dependent SLE, in which participants may play an active role exposure to SLE[39,40]. Different genetic influences, including a higher heritability, are reported for dependent SLE compared to independent SLE[38,41–44], suggesting that whereas GxE driven by independent SLE is likely to reflect a genetic moderation of associations between SLE and depression, GxE driven by dependent SLE may result from a genetic mediation of the association through genetically driven personality or behavioural traits. To test this, we analysed dependent and independent SLE scores separately in Generation Scotland (GS).

Stress contributes to many human conditions, with evidence of genetic vulnerability to the effect of SLE[45]. Therefore, genetic stress-response factors in MDD may also underlie the aetiology of other stress-linked disorders with which MDD is often comorbid[46,47] (e.g., cardiovascular diseases[48], diabetes[49], chronic pain[50] and inflammation[51]). We tested the hypothesis that pleiotropy and shared aetiology between mental and physical health conditions may be due in part to genetic variants underlying SLE effects in depression.

In this study, we conduct GWEIS of depressive symptoms incorporating data on SLE in two independent UK-based cohorts. We aimed to: (i) identify loci associated with depressive symptoms through genetic response to SLE; (ii) study dependent and independent SLE to support a contribution of genetically mediated exposure to stress; (iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone; and (iv) test for a significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes.

Materials and methods

The core workflow of this study is summarised in Fig. 1.

Cohort descriptions

GS

GS is a family-based population cohort representative of the Scottish population[52]. At baseline, blood and salivary DNA samples were collected, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. Genome-wide genotype data were generated using the Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry[53]. The procedures and details for DNA extraction and genotyping have been extensively described elsewhere[54,55]. In total, 21,525 participants were re-contacted to participate in a follow-up mental health study (Stratifying Resilience and Depression Longitudinally, STRADL), of which 8541 participants responded providing updated measures in psychiatric symptoms and SLE through self-reported mental health questionnaires[56]. Samples were excluded if: they were duplicate samples, had diagnoses of bipolar disorder, no SLE data (non-respondents), were population outliers (mainly non-Caucasians and Italian ancestry subgroup), had sex mismatches or were missing >2% of genotypes. Single nucleotide polymorphisms (SNPs) were excluded if: missing >2% of genotypes, Hardy–Weinberg equilibrium test $p < 1 \times 10^{-6}$, or minor allele frequency <1%. Further details of the GS and STRADL cohort are available elsewhere[52,56–58]. All components of GS and STRADL obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the NHS (reference 05/s1401/89). After quality control, individuals were filtered by degree of relatedness (pi-hat < 0.05), maximising retention of those individuals reporting a higher number of SLE. The final dataset comprised data on 4919 unrelated individuals (1929 men; 2990 women) and 560,351 SNPs.

Independent GS datasets

Additional datasets for a range of stress-linked medical conditions and personality traits were created from GS (*N*

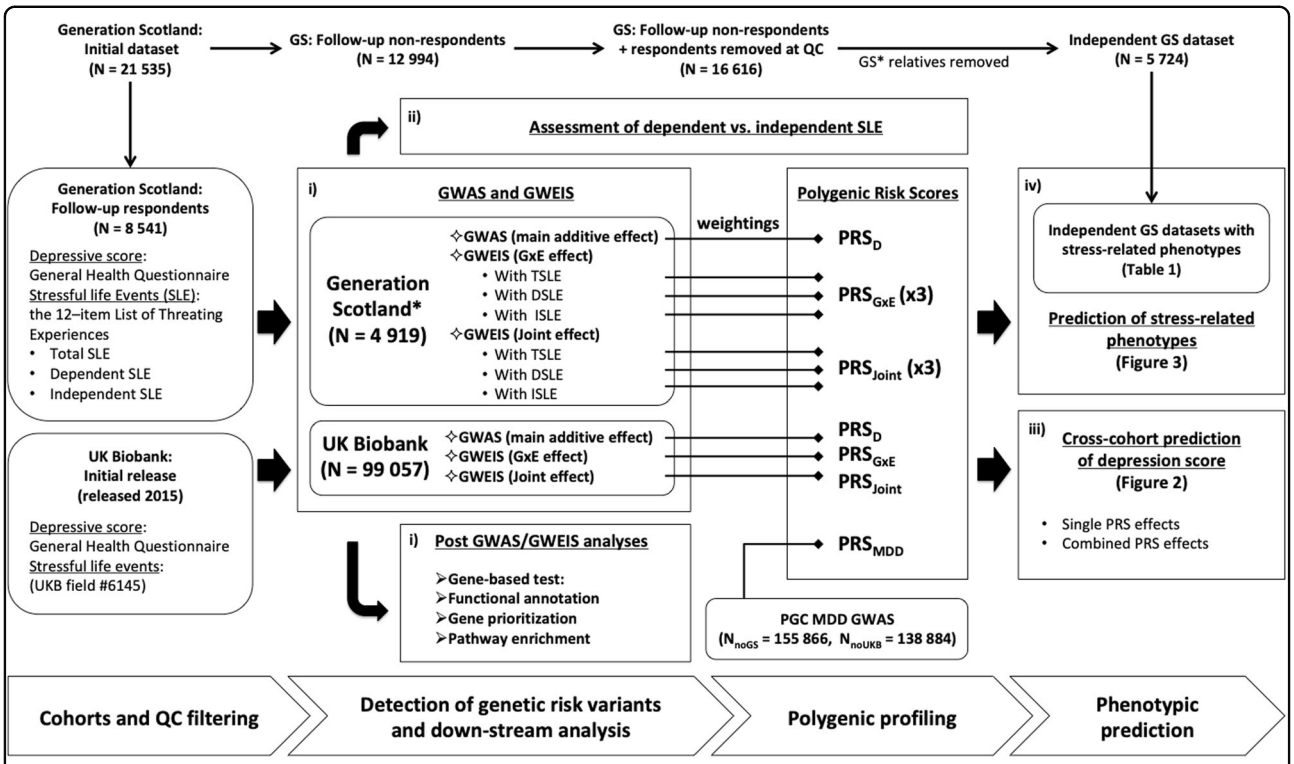


Fig. 1 Study flowchart. Overview of the analyses conducted in this study: (i) identify loci associated with depressive symptoms through genetic response to SLE; (ii) test whether results of studying dependent and independent SLE support a contribution of genetically mediated exposure to stress; (iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone and (iv) test whether there is significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes. Two core cohorts are used, Generation Scotland (GS) and UK Biobank (UKB). Summary statistics from genome-wide association studies (GWAS) and genome-wide by environment interaction studies (GWEIS) are used to generate polygenic risk scores (PRSs). Summary statistics from Psychiatric Genetic Consortium (PGC) Major Depressive Disorder (MDD) GWAS are also used to generate PRS (PRS_{MDD}). PRS weighted by: additive effects (PRS_D and PRS_{MDD}), GxE effects (PRS_{GxE}) and joint effects (the combined additive and GxE effect; PRS_{Joint}), are used for phenotypic prediction. TSLE stands for total number of SLE reported. DSLE stands for SLE dependent on an individual's own behaviour. Conversely, ISLE stands for independent SLE. *N* stands for sample size. N_{noGS} stands for sample size with GS individuals removed. N_{noUKB} stands for sample size with UKB individuals removed

= 21,525) excluding respondents and their relatives (N = 5724). Following the same quality control criteria detailed above, we maximised unrelated non-respondents for retention of cases, or proxy cases (see below), to maximise the information available for each phenotype. This resulted in independent datasets with unrelated individuals for each trait. Differences between respondents and non-respondents are noted in the figure legend of Table 1.

UK Biobank (UKB)

This study used data from 99,057 unrelated individuals (47,558 men; 51,499 women) from the initial release of UKB genotyped data (released 2015; under UKB project 4844). Briefly, participants were removed based on UKB genomic analysis exclusion, non-white British ancestry, high missingness, genetic relatedness (kinship coefficient > 0.0442), QC failure in UK BiLEVE study and gender mismatch. GS participants and their relatives were excluded and GS SNPs imputed to a reference set combining the UK10K haplotype and 1000 Genomes Phase 3 reference panels[59]. After quality control, 1,009,208 SNPs remained. UKB received ethical approval from the NHS National Research Ethics Service North West (reference: 11/NW/0382). Further details on UKB cohort description, genotyping, imputation and quality control are available elsewhere[60–62].

All participants provided informed consent.

Phenotype assessment

SLEs

GS participants reported SLE experienced over the preceding 6 months through a self-reported brief life events questionnaire based on the 12-item list of threatening experiences[39,63,64] (Supplementary Table 1a). The total number of SLE reported (TSLE) consisted of the number of 'yes' responses. TSLE were subdivided into SLE

Table 1 GS samples with stress-related phenotypes

Trait	*N*	Males/females	*N* SNPs	*N* Cases	*N* Controls
Alzheimer (R)	3377	1475/1903	560,622	655	2722
Asthma	3390	1500/1890	560,569	555	2835
Asthma (R)	3375	1470/1905	560,432	910	2465
Bowel cancer (R)	3386	1495/1891	560,630	672	2714
Breast cancer	3388	1486/1902	560,611	83	3305
Breast cancer (R)	3386	1482/1904	560,579	564	2822
Chronic obstructive pulmonary disease	3387	1496/1891	560,591	73	3314
Chronic obstructive pulmonary disease (R)	3387	1474/1913	560,620	553	2834
Depression	3385	1495/1890	560,584	483	2902
Depression (R)	3382	1506/1876	560,514	731	2651
Diabetes	3388	1497/1891	560,469	185	3203
Diabetes (R)	3389	1481/1908	560,584	1144	2245
Heart disease	3392	1504/1888	560,526	212	3180
Heart disease (R)	3377	1483/1894	560,479	2254	1123
High blood pressure	3402	1501/1901	560,508	729	2673
High blood pressure (R)	3372	1464/1908	560,569	1901	1471
Hip fracture (R)	3388	1489/1899	560,572	421	2967
Lung cancer (R)	3379	1492/1887	560,600	798	2581
Osteoarthritis	3395	1486/1909	560,640	411	2984
Osteoarthritis (R)	3383	1466/1917	560,516	961	2422
Parkinson (R)	3388	1488/1900	560,590	236	3152
Prostate cancer (R)	3381	1495/1886	560,570	329	3052
Rheumatoid arthritis	3387	1490/1897	560,618	93	3294
Rheumatoid arthritis (R)	3380	1487/1893	560,543	765	2615
Stroke	3387	1492/1895	560,613	81	3306
Stroke (R)	3385	1463/1922	560,478	1506	1879
Neuroticism[a]	3421	1521/1900	560,484	-	-
Extraversion[a]	3420	1520/1900	560,476	-	-
Schizotypal personality[a]	2386	1065/1321	560,369	-	-
Mood disorder[a]	2307	1040/1267	560,318	-	-

Samples were maximised for retention of cases to maximise the information available for each trait. There was no preferential selection of relatives in pairs for quantitative phenotypes, in order to retain the underlying distribution. All individuals involved in the datasets listed above were non-respondents to the GS follow-up study. Compared with individuals included at GS GWEIS (respondents in GS follow-up), non-respondents were significantly: younger, from more socioeconomically deprived areas, generally less healthier and wealthier. Non-respondents were more likely to smoke, and less likely to drink alcohol, although they consumed more units per week, compared with respondents. At GS baseline, non-respondents experienced more psychological distress and reported higher scores in symptoms of GHQ-depression and GHQ-anxiety than respondents[56]

The total target sample size (*N*), number of males and females in *N*, number of SNPs (*N* SNPs) in target sample size *N*: the number of SNPs used as predictors after clumping step range between 90,650 and 91,000. The number of cases and controls in the independent target sample is indicated for binary phenotypes only. Samples were mapping by proxy approach was used (i.e., where first-degree relatives of individuals with the disease were considered proxy cases and included into the group of cases) are indicated by (R)

GS Generation Scotland, *GWEIS* genome-wide by environment interaction studies, *GHQ* General Health Questionnaire

[a]Assessed through self-reported questionnaires

potentially dependent or secondary to an individual's own behaviour (DSLE, questions 6–11 in Supplementary Table 1a), and independent SLE (ISLE, questions 1–5 in Supplementary Table 1a; pregnancy item removed) following Brugha et al.[39,40]. Thus, three SLE measures (TSLE, DSLE and ISLE) were constructed for GS. UKB

participants were screened for 'illness, injury, bereavement and stress" (Supplementary Table 1b) over the previous 2 years using six items included in the UKB Touchscreen questionnaire. A score reflecting SLE reported in UKB ($TSLE_{UKB}$) was constructed by summing the number of 'yes' responses.

Psychological assessment

GS participants reported whether their current mental state over the preceding 2 weeks differed from their typical state using a self-administered 28-item scaled version of the General Health Questionnaire (GHQ)[65–67]. Participants rated the degree and severity of their current symptoms with a four-point Likert scale (following Goldberg et al.[67]). A final log-transformed GHQ was used to detect altered psychopathology and thus, assess depressive symptoms as results of SLE. In UKB participants, current depressive symptoms over the preceding 2 weeks were evaluated using four psychometric screening items (Supplementary Table 2), including two validated and reliable questions for screening depression[68], from the Patient Health Questionnaire (PHQ) validated to screen mental illness[69,70]. Each question was rated in a four-point Likert scale to assess impairment/severity of symptoms. Due to its skewed distribution, a four-point PHQ score was formed from PHQ (0 = 0; 1 = 1–2; 2 = 3–5; 3 = 6 or more) to create a more normal distribution.

Stress-related traits

Targeted GS stress-related phenotypes and sample sizes are shown in Table 1 and detailed elsewhere[52]. These conditions were selected from literature review based on previous evidence of a link with stress[45] (see also Supplementary Material: third section). Furthermore, we created additional independent samples using mapping by proxy, where individuals with a self-reported first-degree relative with a selected phenotype were included as proxy cases. This approach provides greater power to detect susceptibility variants in traits with low prevalence[71].

Statistical analyses

SNP-heritability and genetic correlation

A restricted maximum likelihood approach was applied to estimate SNP-heritability (h^2_{SNP}) of depressive symptoms and self-reported SLE measures, and within samples bivariate genetic correlation between depressive symptoms and SLE measures using GCTA[72].

GWAS analyses

GWAS were conducted in PLINK[73]. In GS, age, sex and 20 principal components (PCs) were fitted as covariates. In UKB, age, sex and 15 PCs recommended by UKB were fitted as covariates. The genome-wide significance threshold was $p = 5 \times 10^{-8}$.

GWEIS analyses

GWEIS were conducted on GHQ (the dependent variable) for TSLE, DSLE and ISLE in GS and on PHQ for $TSLE_{UKB}$ in UKB fitting the same covariates detailed above to reduce error variance. GWEIS were conducted using an R plugin for PLINK[73] developed by Almli et al.[74] (https://epstein-software.github.io/robust-joint-interaction). This method implements a robust test that jointly considers SNP and SNP–environment interaction effects from a full model ($Y \sim \beta_0 + \beta SNP + \beta SLE + \beta SNP{\times}SLE + \beta Covariates$) against a null model where both the SNP and SNP×SLE effects equal 0, to assess the joint effect (the combined additive main and GxE genetic effect at a SNP) using a nonlinear statistical approach that applies Huber–White estimates of variance to correct possible inflation due to heteroscedasticity (unequal variances across exposure levels). This robust test should reduce confounding due to differences in variance induced by covariate interaction effects if present[75]. Additional code was added (courtesy of Prof. Michael Epstein;[74] Supplementary Material) to generate beta-coefficients and the p-value of the GxE term alone. In UKB, correcting for 1,009,208 SNPs and one exposure, we established a Bonferroni-adjusted threshold for significance at $p = 2.47 \times 10^{-8}$ for both joint and GxE effects. In GS, correcting for 560,351 SNPs and three measures of SLE we established a genome-wide significance threshold of $p = 2.97 \times 10^{-8}$.

Post-GWAS/GWEIS analyses

GWAS and GWEIS summary statistics were analysed using FUMA[76] including: gene-based tests, functional annotation, gene prioritisation and pathway enrichment (Supplementary Material).

Polygenic profiling and prediction

Polygenic risk scores (PRSs) weighting by GxE effects (PRS_{GxE}) were generated using PRSice-2[77] (Supplementary Material) in GS using GxE effects from UKB-GWEIS. In UKB, PRS_{GxE} were constructed using GxE effects from all three GS-GWEIS (TSLE, DSLE and ISLE as exposures) independently. PRS were also weighted in both samples using either UKB-GWAS or GS-GWAS statistics (PRS_D), and summary statistics from Psychiatric Genetic Consortium (PGC) MDD-GWAS (released 2016; PRS_{MDD}) that excluded GS and UKB individuals when required ($N_{noGS} = 155{,}866$; $N_{noUKB} = 138{,}884$). Furthermore, we calculated PRS weighted by the joint effects (the combined additive main and GxE genetic effects; PRS_{Joint}) from either the UKB-GWEIS or GS-GWEIS. PRS predictions of depressive symptoms were permuted 10,000 times. Multiple regression models fitting PRS_{GxE} and PRS_{MDD}, and both PRS_{GxE} and PRS_D were tested. All models were adjusted by same covariates used in GWAS/

GWEIS. Null models were estimated from the direct effects of covariates alone. The predictive improvement of combining PRS_{GxE} and PRS_{MDD}/PRS_D effects over PRS_{MDD}/PRS_D effects alone was tested for significance using the likelihood ratio test (LRT).

Prediction of PRS_D, PRS_{GxE} and PRS_{Joint} of stress-linked traits were adjusted by age, sex and 20 PCs; and permuted 10,000 times. Empirical-*p*-values after permutations were further adjusted by false discovery rate (FDR, conservative threshold at Empirical-$p = 6.16 \times 10^{-3}$). The predictive improvement of fitting PRS_{GxE} combined with PRS_D and covariates over prediction of a phenotype using the PRS_D effect alone with covariates was assessed using LRT, and *LRT-p*-values adjusted by FDR (conservative threshold at $LRT\text{-}p = 8.35 \times 10^{-4}$).

Results

Phenotypic and genetic correlations

Depressive symptom scores and SLE measures were positively correlated in both UKB ($r^2 = 0.22$, $p < 2.2 \times 10^{-16}$) and GS (TSLE-$r^2 = 0.21$, $p = 1.69 \times 10^{-52}$; DSLE-$r^2 = 0.21$, $p = 8.59 \times 10^{-51}$; ISLE-$r^2 = 0.17$, $p = 2.33 \times 10^{-33}$). Significant bivariate genetic correlation between depression and SLE scores was identified in UKB (rG = 0.72; $p < 1 \times 10^{-5}$, $N = 50{,}000$), but not in GS (rG = 1, $p = 0.056$, $N = 4919$; Supplementary Table 3a).

SNP-heritability (h^2_{SNP})

In UKB, a significant h^2_{SNP} of PHQ was identified ($h^2_{SNP} = 0.090$; $p < 0.001$; $N = 99{,}057$). This estimate remained significant after adjusting by $TSLE_{UKB}$ effect ($h^2_{SNP} = 0.079$; $p < 0.001$), suggesting a genetic contribution unique to depressive symptoms. The h^2_{SNP} of $TSLE_{UKB}$ was also significant ($h^2_{SNP} = 0.040$, $p < 0.001$; Supplementary Table 3b). In GS, h^2_{SNP} was not significant for GHQ ($h^2_{SNP\,=}\,0.071$, $p = 0.165$; $N = 4919$). However, in an *ad hoc* estimation from the baseline sample of 6751 unrelated GS participants (details in Supplementary Table 3b) we detected a significant h^2_{SNP} for GHQ ($h^2_{SNP} = 0.135$; $p < 5.15 \times 10^{-3}$), suggesting that the power to estimate h^2_{SNP} in GS may be limited by sample size. Estimates were not significant for either TSLE ($h^2_{SNP} = 0.061$, $p = 0.189$; Supplementary Table 3b) or ISLE ($h^2_{SNP} = 0.000$, $p = 0.5$), but h^2_{SNP} was significant for DSLE ($h^2_{SNP} = 0.131$, $p = 0.029$), supporting a potential genetic mediation and gene–environment correlation.

GWAS of depressive symptoms

No genome-wide significant SNPs were detected by GWAS in either cohort. Top findings ($p < 1 \times 10^{-5}$) are summarised in Supplementary Table 4. Manhattan and QQ plots are shown in Supplementary Figures 1-4. There was no evidence of genomic inflation (all $\lambda_{1000} < 1.01$).

Post-GWAS analyses

Gene-based test identified six genes associated with PHQ using the UKB-GWAS statistics at genome-wide significance (Bonferroni-corrected $p = 2.77 \times 10^{-6}$; *DCC*, $p = 7.53 \times 10^{-8}$; *ACSS3*, $p = 6.51 \times 10^{-7}$; *DRD2*, $p = 6.55 \times 10^{-7}$; *STAG1*, $p = 1.63 \times 10^{-6}$; *FOXP2*, $p = 2.09 \times 10^{-6}$; *KYNU*, $p = 2.24 \times 10^{-6}$; Supplementary Figure 8). Prioritised genes based on position, expression quantitative trait loci (eQTL) and chromatin interaction mapping are detailed in Supplementary Table 5. No genes were detected in GS-GWAS gene-based test (Supplementary Figures 9). No tissue-specific enrichment was detected from GWAS in either cohort. Significant gene-sets and GWAS catalogue associations for UKB-GWAS are reported in Supplementary Table 6. These included the *biological process*: positive regulation of long-term synaptic potentiation, and *GWAS catalogue associations*: brain structure, schizophrenia, response to amphetamines, age-related cataracts (age at onset), fibrinogen, acne (severe), fibrinogen levels and educational attainment; all adjusted-$p < 0.01$. There was no significant gene-set enrichment from GS-GWAS.

GWEIS of depressive symptoms

Manhattan and QQ plots are shown in Supplementary Figures 1-4. There was no evidence of GWEIS inflation for either UKB or GS (all $\lambda_{1000} < 1.01$). No genome-wide significant GWEIS associations were detected for SLE in UKB. GS-GWEIS using TSLE identified a significant GxE effect ($p < 2.97 \times 10^{-8}$) at an intragenic SNP on chromosome 11 (rs12789145, $p = 4.95 \times 10^{-9}$, $\beta = 0.06$, closest gene: *PIWIL4*; Supplementary Figure 5), and using DSLE at an intronic SNP in *ZCCHC2* on chromosome 18 (rs17070072, $p = 1.46 \times 10^{-8}$, $\beta = -0.08$; Supplementary Figure 6). In their corresponding joint effect tests, both rs12789145 ($p = 2.77 \times 10^{-8}$) and rs17070072 $p = 1.96 \times 10^{-8}$) were significant. GWEIS for joint effect using DSLE identified two further significant SNPs on chromosome 9 (rs12000047, $p = 2.00 \times 10^{-8}$, $\beta = -0.23$; rs12005200, $p = 2.09 \times 10^{-8}$, $\beta = -0.23$, LD $r^2 > 0.8$, closest gene: *CYLC2*; Supplementary Figure 7). None of these associations replicated in UKB ($p > 0.05$), although the effect direction was consistent between cohorts for the SNP close to *PIWL1* and SNPs at *CYLC2*. No SNP achieved genome-wide significant association in the GS-GWEIS using ISLE as exposure. Top GWEIS results ($p < 1 \times 10^{-5}$) are summarised in Supplementary Tables 7-10.

Post-GWEIS analyses: gene-based tests

All results are shown in Supplementary Figures 10-17. Two genes were associated with PHQ using the joint effect from the UKB-GWEIS (*ACSS3* $p = 1.61 \times 10^{-6}$; *PHF2*, $p = 2.28 \times 10^{-6}$; Supplementary Figure 11). *ACSS3* was previously identified using the additive main effects,

whereas *PHF2* was only significantly associated using the joint effects. Gene-based tests identified *MTNR1B* as significantly associated with GHQ on the GS-GWEIS using DSLE in both GxE ($p = 1.53 \times 10^{-6}$) and joint effects ($p = 2.38 \times 10^{-6}$; Supplementary Figures 14-15).

Post-GWEIS analyses: tissue enrichment

We prioritised genes based on position, eQTL and chromatin interaction mapping in brain tissues and regions. In UKB, prioritised genes using GxE effects were enriched for upregulated differentially expressed genes from adrenal gland (adjusted-$p = 3.58 \times 10^{-2}$). Using joint effects, prioritised genes were enriched on upregulated differentially expressed genes from artery tibial (adjusted-$p = 4.34 \times 10^{-2}$). In GS, prioritised genes were enriched: in upregulated differentially expressed genes from artery coronary (adjusted-$p = 4.55 \times 10^{-2}$) using GxE effects with DSLE; in downregulated differentially expressed genes from artery aorta tissue (adjusted-$p = 4.71 \times 10^{-2}$) using GxE effects with ISLE; in upregulated differentially expressed genes from artery coronary (adjusted-$p = 5.97 \times 10^{-3}$, adjusted-$p = 9.57 \times 10^{-3}$) and artery tibial (adjusted-$p = 1.05 \times 10^{-2}$, adjusted-$p = 1.55 \times 10^{-2}$) tissues using joint effects with both TSLE and DSLE; and in downregulated differentially expressed genes from lung tissue (adjusted-$p = 3.98 \times 10^{-2}$) and in up- and downregulated differentially expressed genes from the spleen (adjusted-$p = 4.71 \times 10^{-2}$) using joint effects with ISLE. There was no enrichment using GxE effect with TSLE.

Post-GWEIS analyses: gene-sets enrichment

Significant gene-sets and GWAS catalogue hits from GWEIS are detailed in Supplementary Tables 11-14, including for UKB *Biocarta*: GPCR pathway; *Reactome*: opioid signalling, neurotransmitter receptor binding and downstream transmission in the postsynaptic cell, transmission across chemical synapses, gastrin CREB signalling pathway via PKC and MAPK; *GWAS catalogue*: post bronchodilator FEV1/FVC ratio, migraine and body mass index. In GS, enrichment was seen using TSLE and DLSE for *GWAS catalogue*: age-related macular degeneration, myopia, urate levels and Heschl's gyrus morphology; and using ISLE for *biological process*: regulation of transporter activity. All adjusted-$p < 0.01$.

Cross-cohort prediction

In GS, PRS_D weighted by the UKB-GWAS of PHQ significantly explained 0.56% of GHQ variance (Empirical-$p < 1.10^{-4}$), similar to PRS_{MDD} weighted by PGC MDD-GWAS ($R^2 = 0.78\%$, Empirical-$p < 1.10^{-4}$). PRS_{GxE} weighted by the UKB-GWEIS GxE effects explained 0.15% of GHQ variance (Empirical-$p = 0.03$, Supplementary Table 15). PRS_{GxE} fitted jointly with PRS_{MDD} significantly improved prediction of GHQ ($R^2 = 0.93\%$, model $p = 6.12 \times 10^{-11}$; predictive improvement of 19%, *LRT*-$p = 5.91 \times 10^{-3}$) compared with PRS_{MDD} alone. Similar to PRS_{GxE} with PRS_D ($R^2 = 0.69\%$, model $p = 2.72 \times 10^{-8}$; predictive improvement of 23%, *LRT*-$p = 0.01$). PRS_{Joint} weighted by the UKB-GWEIS also predicted GHQ ($R^2 = 0.58\%$, Empirical-$p < 1.10^{-4}$), although

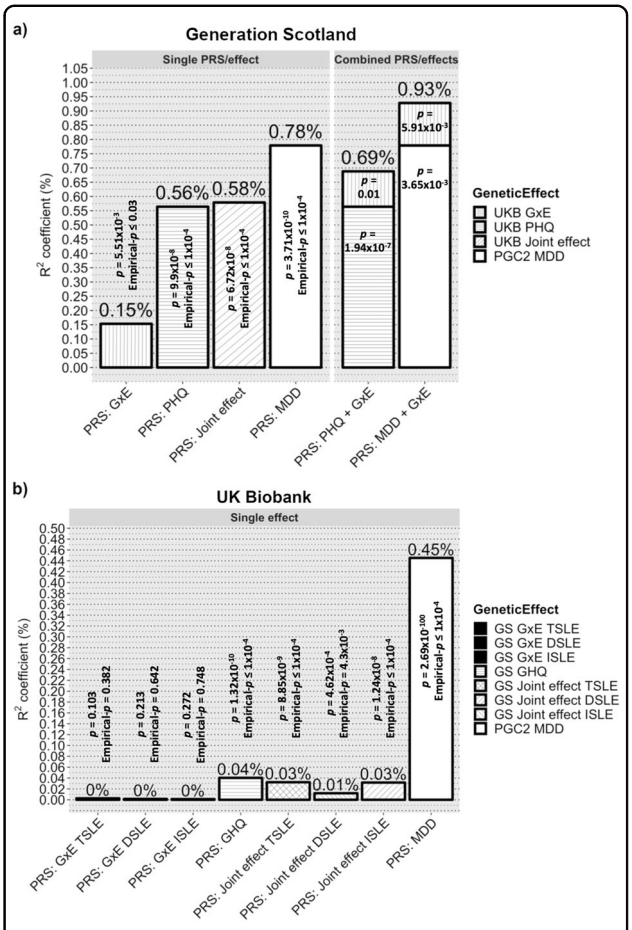


Fig. 2 Prediction of depression scores by PRS_{GxE}, PRS_D, PRS_{MDD} and PRS_{Joint}. Variance of depression score explained by PRS_{GxE} PRS_D, PRS_{MDD} and PRS_{Joint} as single effect; and combining both PRS_D and PRS_{MDD} with PRS_{GxE} in single models. Prediction was conducted using **a** Generation Scotland (GS) and **b** UK Biobank (UKB) as target sample. PRS_{GxE} were weighted by cross- sample genome-wide by environment interaction studies (GWEIS) using GxE effect. PRS_D were weighted by cross-sample genome-wide association studies (GWAS) of depressive symptoms effect. PRS_{MDD} was weighted by Psychiatric Genetic Consortium (PGC) Major Depressive Disorder (MDD)-GWAS summary statistics. PRS_{Joint} were weighted by cross-sample GWEIS using joint effect. A nominally significant gain in variance explained of General Health Questionnaire (GHQ) of about 23% was seen in GS when PRS_{GxE} was incorporated into a multiple regression model along with PRS_D; and of about 19% when PRS_{GxE} was incorporated into a multiple regression model along with PRS_{MDD}. Such a gain was not seen in UKB, but it must be noted that both PRS_D and PRS_{MDD} also explains much less variance of PHQ in UKB than of GHQ in GS. Also note, a noticeably reduction of variance explained by PRS_{Joint} compared with combined polygenic risk scores (PRS)/effects

the variance explained was significantly reduced compared with the model fitting PRS_{GxE} and PRS_D together ($LRT\text{-}p = 4.69 \times 10^{-7}$), suggesting that additive and GxE effects should be modelled independently for polygenic approaches (Fig. 2a).

In UKB (Fig. 2b), both PRS_D weighted by the GS-GWAS of GHQ and PRS_{MDD} significantly explained 0.04 and 0.45% of PHQ variance, respectively (both Empirical-$p < 1.10^{-4}$; Supplementary Table 15). PRS_{GxE} derived from the GS-GWEIS GxE effect did not significantly predicted PHQ (TSLE-PRS_{GxE} Empirical-$p = 0.382$; DSLE-PRS_{GxE} Empirical-$p = 0.642$; ISLE-PRS_{GxE} Empirical-$p = 0.748$). Predictive improvements using the PRS_{GxE} effect fitted jointly with PRS_{MDD} or PRS_D were not significant (all $LRT\text{-}p > 0.08$). PRS_{Joint} significantly predicted PHQ (TSLE-PRS_{Joint}: $R^2 = 0.032\%$, Empirical-$p < 1.10^{-4}$; DSLE-PRS_{Joint}: $R^2 = 0.012\%$, Empirical-$p = 4.3 \times 10^{-3}$; ISLE-PRS_{Joint}: $R^2 = 0.032\%$, Empirical-$p < 1.10^{-4}$), although the variance explained was significantly reduced compared with the models fitting PRS_{GxE} and PRS_D together (all $LRT\text{-}p < 1.48 \times 10^{-3}$).

Prediction of stress-related traits

Prediction of stress-related traits in independent samples using PRS_D, PRS_{GxE} and PRS_{Joint} are summarised in Fig. 3a and Supplementary Table 16. Significant trait prediction after FDR adjustment (Empirical-$p < 6.16 \times 10^{-3}$, FDR-adjusted Empirical-$p < 0.05$) using both UKB and GS PRS_D was seen for: depression status, neuroticism and schizotypal personality. PRS_{GxE} weighted by the GS-GWEIS GxE effect using ISLE significantly predicted depression status mapped by proxy (Empirical-$p = 7.00 \times 10^{-4}$, FDR-adjusted Empirical-$p = 9.54 \times 10^{-3}$).

Nominally significant predictive improvements ($LRT\text{-}p < 0.05$) of fitting PRS_{GxE}, over the PRS_D effect alone, using summary statistics generated from both UKB and GS were detected for schizotypal personality, heart diseases and chronic obstructive pulmonary disease (COPD) by proxy (Fig. 3b). PRS_{GxE} weighted by GS-GWEIS GxE effect using ISLE significantly improved prediction over PRS_D effect alone of depression status mapped by proxy after FDR adjustment ($LRT\text{-}p = 1.96 \times 10^{-4}$, *FDR-adjusted* $LRT\text{-}p = 2.35 \times 10^{-2}$).

Discussion

This study performs GWAS and incorporates data on recent adult SLEs into GWEIS of depressive symptoms, identifies new loci and candidate genes for the modulation of genetic response to SLE; and provides insights to help disentangle the underlying aetiological mechanisms increasing the genetic liability through SLE to both depressive symptoms and stress-related traits.

SNP-heritability of depressive symptoms (h^2_{SNP} = 9–13%), were slightly higher than previous estimates from African-American populations[34], and over a third larger than estimates in MDD from European samples[78]. h^2_{SNP} for PHQ in UKB (9.0%) remained significant after adjusting for SLE (7.9%). Thus, although some genetic contributions may be partially shared between depressive symptoms and reporting of SLE, there is still a relatively large genetic contribution unique to depressive symptoms. Significant h^2_{SNP} of DSLE in GS (13%) and $TSLE_{UKB}$ in UKB (4%), which is mainly composed of dependent SLE items, were detected similar to previous studies (8 and 29%)[34,42]. Conversely, there was no evidence for heritability of independent SLE. A significant bivariate genetic correlation between depressive symptoms and SLE (rG = 0.72) was detected in UKB after adjusting for covariates, suggesting that there are shared common variants underlying self-reported depressive symptoms and SLE. This bivariate genetic correlation was smaller than that estimated from African-American populations (rG = 0.97; $p = 0.04$; $N = 7179$)[34]. Genetic correlations between SLE measures and GHQ were not significant in GS ($N = 4919$; rG = 1; all $p > 0.056$), perhaps due to a lack of power in this smaller sample.

Post-GWAS gene-based tests detected six genes significantly associated with PHQ (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*). Previous studies have implicated these genes in liability to depression (see Supplementary Table 17), and three of them are genome-wide significant in gene-based tests from the latest meta-analysis of major depression that includes UKB (*DCC*, $p = 2.57 \times 10^{-14}$; *DRD2*, $p = 5.35 \times 10^{-14}$; and *KYNU*, $p = 2.38 \times 10^{-6}$; $N = 807{,}553$)[79]. This supports the implementation of quantitative measures such as PHQ to detect genes underlying lifetime depression status[80]. For example, significant gene ontology analysis of the UKB-GWAS identified enrichment for positive regulation of long-term synaptic potentiation, and for previous GWAS findings of brain structure[81], schizophrenia[82] and response to amphetamines[83].

The key element of this study was to conduct GWEIS of depressive symptoms and recent SLE. We identified two loci with significant GxE effect in GS. However, none of these associations replicated in UKB ($p > 0.05$). The strongest association was using TSLE at 53-kb downstream of *PIWIL4* (rs12789145). *PIWIL4* is brain expressed and involved in chromatin modification[84], suggesting it may moderate the effects of stress on depression. It encodes HIWI2, a protein thought to regulate OTX2, that is critical for the development of forebrain and for coordinating critical periods of plasticity disrupting the integration of cortical circuits[85,86]. Indeed, an intronic SNP in *PIWIL4* was identified as the strongest GxE signal in attention deficit hyperactivity disorder using mother's warmth as environmental exposure[87]. The other significant GxE identified in our study was in *ZCCHC2* using

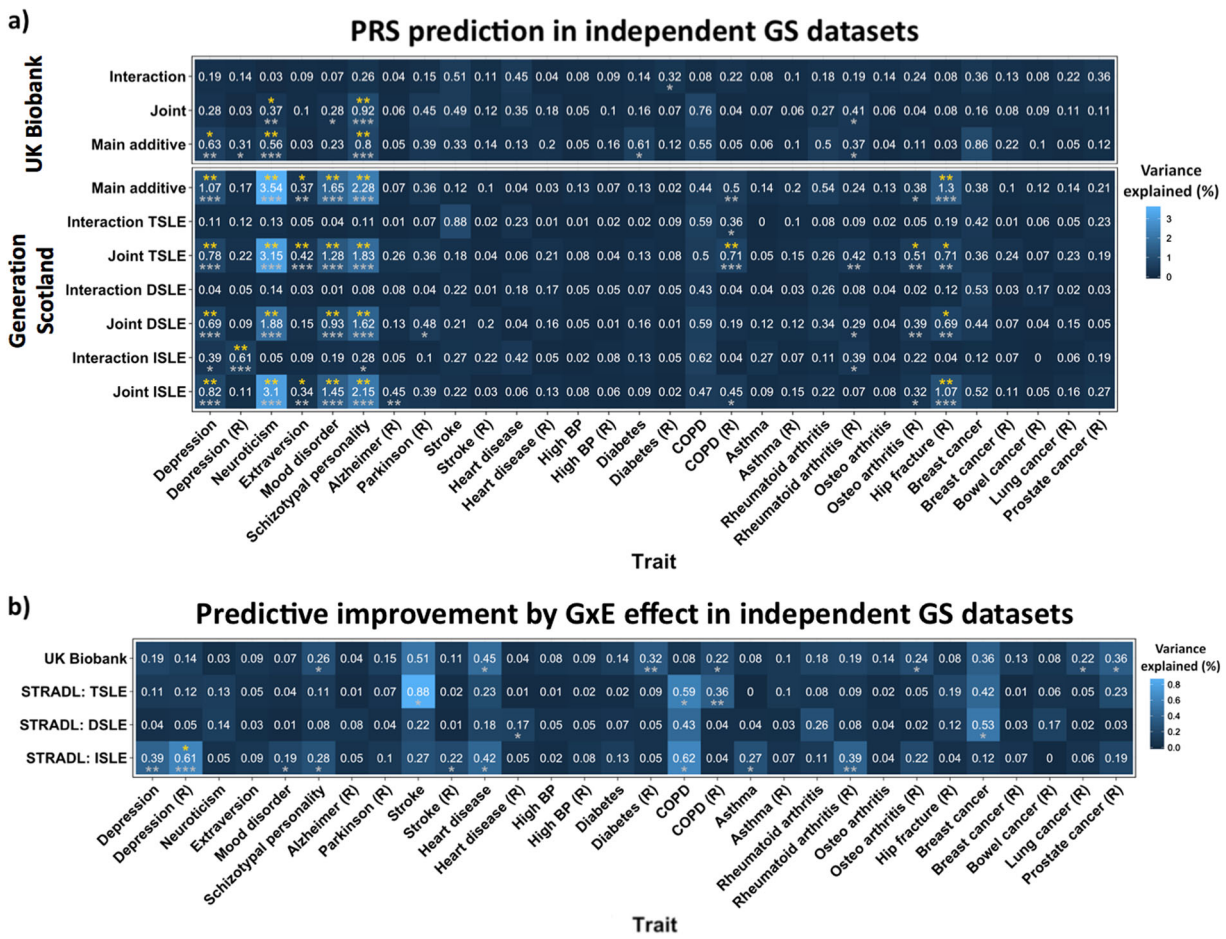


Fig. 3 Polygenic risk score (PRS) prediction in independent Generation Scotland (GS) datasets. a Heatmap illustrating PRS prediction of a wide range of traits from GS listed in the *x* axis (Table 1). (R) refers to traits using mapping by proxy approach (i.e., where first-degree relatives of individuals with the disease are considered proxy cases and included into the group of cases). *Y* axis shows the discovery sample and the effect used to weight PRS. Numbers in cells indicate the % of variance explained, also represented by colour scale. Significance is represented by asterixes according to the following significance codes: $^{**}p < 0.01$; $^{*}p < 0.05$; in grey Empirical-*p*-values after permutation (10,000 times) and in yellow FDR-adjusted Empirical-*p*-values. **b** Predictive improvement by GxE effect in independent GS datasets. Heatmap illustrating the predictive improvement as a result of incorporating PRS_{GxE} into a multiple model along with PRS_D and covariates (full model), over a model fitting PRS_D alone with covariates (null model); predicting a wide range of traits from GS listed in the *x* axis (Table 1). Covariates: age, sex and 20 PCs. (R) refers to traits using mapping by proxy approach (i.e., where first-degree relatives of individuals with the disease are consider proxy cases and included into the group of cases). PRS_{GxE} are weighted by genome-wide by environment interaction studies (GWEIS) using GxE effects. PRS_D were weighted by the genome-wide association studies (GWAS) of depressive symptoms additive main effects. The *y* axis shows the discovery sample used to weight PRS. Numbers in cells indicate the % of variance explained by the PRS_{GxE}, also represented by colour scale. Notice that those correspond to the PRS_{GxE} predictions in Fig. 3a when PRS_{GxE} are weighted by GxE effects. Significance was tested by likelihood ratio tests (LRT): full model including $PRS_D + PRS_{GxE}$ vs. null model with PRS_D alone (covariates adjusted). Significance is represented by asterixes according to the following significance codes: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; in grey *LRT-p*-values and in yellow FDR-adjusted *LRT-p*-values

DSLE. This zinc-finger protein is expressed in blood CD4 + T cells and is downregulated in individuals with MDD[88] and in those resistant to treatment with citalopram[89]. No GxE effect was seen using ISLE as exposure.

No significant locus or gene with GxE effect was detected in the UKB-GWEIS. Nevertheless, joint effects (the combined additive main and GxE genetic effects) identified two genes significantly associated with PHQ (*ACSS3* and *PHF2*; see Supplementary Table 17). *PHF2* was recently detected as genome-wide significant at the latest meta-analysis of depression[79]. Notably, *PHF2* paralogs have previously been linked with MDD through stress-response in three other studies[90–92]. Joint effects analyses in GS also detected an additional significant association upstream *CYLC2*, a gene nominally associated ($p < 1 \times 10^{-5}$) with obsessive-compulsive disorder and Tourette's syndrome[93]. Gene-based test from the GS-GWEIS identified a significant association with

MTNR1B, a melatonin receptor gene, using DSLE (both GxE and joint effect; Supplementary Table 17). Genes prioritised using GxE effects were enriched in differentially expressed genes from several tissues including the adrenal gland, which releases cortisol into the bloodstream in response to stress, thus playing a key role in the stress-response system, reinforcing a potential role of GxE in stress-related conditions.

The different instruments and sample sizes available make it hard to compare results between cohorts. Whereas GS contains deeper phenotyping measurements of stress and depressive symptoms than UKB, the sample size is much smaller, which may be reflected in the statistical power required to reliably detect GxE effects. Furthermore, the presence and size of GxE effects are dependent on their parameterisation (i.e., the measurement, scale and distribution of the instruments used to test such interaction)[94]. Thus, GxE may be incomparable across GWEIS due to differences in both phenotype assessment and stressors tested. Although our results suggest that both depressive symptom measures are correlated with lifetime depression status, different influences on depressive symptoms from the SLE covered across studies may contribute to lack of stronger replication. Instruments in GS cover a wider range of SLE and are more likely to capture changes in depressive symptoms as consequence of their short-term effects. Conversely, UKB could capture more marked long-term effects, as SLE were captured over 2 years compared with the 6 months in GS. New mental health questionnaires covering a wide range of psychiatric symptoms and SLE in the latest release of UKB data provides the opportunity to create similar measures to GS in the near future. Further replication in independent studies with equivalent instruments is required to validate our GWEIS findings. Despite these limitations and a lack of overlap in the individual genes prioritised from the two GWEIS, replication was seen in the predictive improvement of using PRS_{GxE} derived from the GWEIS GxE effects to predict stress-related phenotypes.

The third aim of this study was to test whether modelling GxE effects could improve predictive genetic models, and thus help to explain deviation from additive models and missing heritability for MDD[95]. Multiple regression models suggested that inclusion of PRS_{GxE} weighted by GxE effects could improve prediction of an individual's depressive symptoms over use of PRS_{MDD} or PRS_{D} weighted by additive effects alone. In GS, we detected a predictive gain of 19% over PRS_{MDD} weighted by PGC MDD-GWAS, and a gain of 23% over PRS_{D} weighted by UKB-GWAS (Fig. 2a). However, these findings did not surpass stringent Bonferroni correction and could not be validated in UKB. This may reflect in the poor predictive power of the PRS generated from the much smaller GS discovery sample. The results show a noticeably reduced prediction using PRS_{Joint} weighted by joint effects, which suggests that the genetic architecture of stress-response is at least partially independent and differs from genetic additive main effects. Overall, our results from multiple regression models suggest that for polygenic approaches main and GxE effects should be modelled independently.

SLE effects are not limited to mental illness[45]. Our final aim was to investigate shared aetiology between GxE for depressive symptoms and stress-related traits. Despite the differences between the respondents and non-respondents (Table 1 legend), a significant improvement was seen in predicting depressive status when mapping by proxy cases using GxE effect from GS-GWEIS with independent SLE (*FDR-adjusted LRT-p* = 0.013), but not with dependent SLE. GxE effects using statistics generated from both discovery samples, despite the differences in measures, nominally improved the phenotypic prediction of schizotypal personality, heart disease and the proxy of COPD (*LRT-p* < 0.05). Other studies have also found evidence supporting a link between stress and depression in these phenotypes (see Supplementary Material for extended review) and suggest, for instance, potential pleiotropy between schizotypal personality and stress-response. Our findings point to a potential genetic component underlying a stress-response-depression-comorbidities link due, at least in part, to shared stress-response mechanisms. A relationship between SLE, depression and coping strategies such as smoking suggests that genetic stress-response may modulate adaptive behaviours such as smoking, fatty diet intake, alcohol consumption and substance abuse. This is discussed further in the Supplementary Material.

In this study, evidence for SNPs with significant GxE effects came primarily from the analyses of dependent SLE and not from independent SLE. This supports a genetic effect on probability of exposure to, or reporting of SLE, endorsing a gene–environment correlation. Chronic stress may influence cognition, decision making and behaviour eventually leading to higher risk taking[96]. These conditions may also increase sensitivity to stress among vulnerable individuals, including those with depression, who also have a higher propensity to report SLE, particularly dependent SLE[38]. A potential reporting bias in dependent SLE may be mediated as well by heritable behavioural, anxiety or psychological traits such as risk taking[42,97]. Furthermore, individuals vulnerable to MDD may behave in a manner that exposes them more frequently to high risk and stressful environments[14]. This complex interplay, reflected in the form of a gene–environment correlation effect, would hinder the interpretation of GxE effects from GWEIS as pure interactions. A mediation of associations between SLE and

depressive symptoms, through genetically driven sensitivity to stress, personality or behavioural traits would support the possibility of subtle genotype-by-genotype (GxG) interactions, or genotype-by-genotype-by-environment (GxGxE) interactions, contributing to depression[98,99]. In contrast, PRS prediction of the stress-related traits: schizotypal personality, heart disease and COPD, was primarily from derived weights using independent SLE, suggesting that a common set of variants moderate the effects of SLE across stress-related traits and that larger sample sizes will be required to detect the individual SNPs contributing to this. Thus, our findings support the inclusion of environmental information into GWAS to enhance the detection of relevant genes. The results of studying dependent and independent SLE support a contribution of genetically mediated exposure to and/or reporting of SLE, perhaps through sensitivity to stress as mediator.

This study emphasises the relevance of GxE in depression and human health in general and provides the basis for future lines of research.

Acknowledgements
See Supplementary Material: acknowledgements.

Author details
[1]Medical Genetics Section, University of Edinburgh, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK. [2]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK. [3]Division of Psychiatry, Deanery of Clinical Sciences, Univ×ersity of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK. [4]Health Informatics Centre, University of Dundee, Dundee, UK. [5]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Conflict of interest
The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (https://doi.org/10.1038/s41398-018-0360-y).

Received: 26 November 2018 Accepted: 10 December 2018
Published online: 17 January 2019

References

1. Plomin, R., Owen, M. & McGuffin, P. The genetic basis of complex human behaviors. *Science* **264**, 1733–1739 (1994).
2. Kendler, K. S. & Eaves, L. J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am. J. Psychiatry* **143**, 279–289 (1986).
3. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).
4. Colodro-Conde, L. et al. A direct test of the diathesis-stress model for depression. *Mol. Psychiatry* **23**, 1590–1596 (2017).
5. Luciano, M. et al. Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. *Intelligence* **38**, 304–313 (2010).
6. Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu. Rev. Public Health* **34**, 119–138 (2013).
7. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
8. Hammen, C. Stress and depression. *Annu. Rev. Clin. Psychol.* **1**, 293–319 (2005).
9. Kessler, R. C. The effects of stressful life events on depression. *Annu. Rev. Psychol.* **48**, 191–214 (1997).
10. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* **156**, 837–841 (1999).
11. Stroud, C. B., Davila, J. & Moyer, A. The relationship between stress and depression in first onsets versus recurrences: a meta-analytic review. *J. Abnorm. Psychol.* **117**, 206–213 (2008).
12. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. *J. Nerv. Ment. Dis.* **186**, 661–669 (1998).
13. Silberg, J., Rutter, M., Neale, M. & Eaves, L. Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *Br. J. Psychiatry* **179**, 116–121 (2001).
14. Kendler, K. S. et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am. J. Psychiatry* **152**, 833–842 (1995).
15. Arnau-Soler, A., Adams, M., Hayward, C. & Thomson, P. Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS One* **13**, e0209160 (2018).
16. Arnau-Soler, A. et al. A validation of the diathesis-stress model for depression in Generation Scotland. *Transl. Psychiatry* (in press).
17. Garantziotis, S. & Schwartz, D. A. Ecogenomics of respiratory diseases of public health significance. *Annu. Rev. Public Health* **31**, 37–51 (2010).
18. Aschard, H. et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int. J. Epidemiol.* **46**, 894–904 (2017).
19. Molfino, N. A. & Coyle, A. J. Gene-environment interactions in chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.* **3**, 491–497 (2008).
20. Polonikov, A. V., Ivanov, V. P. & Solodilova, M. A. Genetic variation of genes for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the importance of gene-gene and gene-environment interactions for disease susceptibility. *J. Hum. Genet.* **54**, 440–449 (2009).
21. Haiman, C. A. et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *N. Engl. J. Med.* **354**, 333–342 (2006).
22. Han, J., Hankinson, S. E., Colditz, G. A. & Hunter, D. J. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *Br. J. Cancer* **91**, 1604–1609 (2004).
23. Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
24. Wang, L., Murk, W. & DeWan, A. T. Genome-wide gene by environment interaction analysis identifies common SNPs at 17q21.2 that are associated with increased body mass index only among asthmatics. *PLoS One* **10**, e0144114 (2015).
25. Siegert, S. et al. Genome-wide investigation of gene-environment interactions in colorectal cancer. *Hum. Genet.* **132**, 219–231 (2013).
26. Gong, J. et al. Genome-wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. *PLoS. Genet.* **12**, e1006296 (2016).
27. Polfus, L. M. et al. Genome-wide association study of gene by smoking interactions in coronary artery calcification. *PLoS One* **8**, e74642 (2013).
28. Duncan, L. E. & Keller, M. C. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* **168**, 1041–1049 (2011).
29. Risch, N. et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* **301**, 2462–2471 (2009).
30. Karg, K., Burmeister, M., Shedden, K. & Sen, S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* **68**, 444–454 (2011).
31. Bleys, D., Luyten, P., Soenens, B. & Claes, S. Gene-environment interactions between stress and 5-HTTLPR in depression: a meta-analytic update. *J. Affect Disord.* **226**, 339–345 (2018).

32. Peyrot, W. J. et al. Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biol. Psychiatry* **84**, 138–147 (2018).
33. Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J. & Gauderman, W. J. Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–119 (2007).
34. Dunn, E. C. et al. Genome-wide association study (GWAS) and genome-wide by environment interaction study (GWEIS) of depressive symptoms in African American and Hispanic/Latina women. *Depress. Anxiety* **33**, 265–280 (2016).
35. Otowa, T. et al. The first pilot genome-wide gene-environment study of depression in the Japanese population. *PLoS One* **11**, e0160823 (2016).
36. Ikeda, M. et al. Genome-wide environment interaction between depressive state and stressful life events. *J. Clin. Psychiatry* **77**, e29–e30 (2016).
37. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* **84**, 309–322 (1977).
38. Clarke, T. et al. Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism [version 1; referees: 3 approved with reservations]. *Wellcome Open Res.* **3**, 11 (2019).
39. Brugha, T., Bebbington, P., Tennant, C. & Hurry, J. The list of threatening experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol. Med.* **15**, 189–194 (1985).
40. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. The assessment of dependence in the study of stressful life events: validation using a twin design. *Psychol. Med.* **29**, 1455–1460 (1999).
41. Plomin, R., Lichtenstein, P., Pedersen, N. L., McClearn, G. E. & Nesselroade, J. R. Genetic influence on life events during the last half of the life span. *Psychol. Aging* **5**, 25–30 (1990).
42. Power, R. A. et al. Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol. Med.* **43**, 1965–1971 (2013).
43. Bemmels, H. R., Burt, S. A., Legrand, L. N., Iacono, W. G. & McGue, M. The heritability of life events: an adolescent twin and adoption study. *Twin. Res. Hum. Genet.* **11**, 257–265 (2008).
44. Boardman, J. D., Alexander, K. B. & Stallings, M. C. Stressful life events and depression among adolescent twin pairs. *Biodemography Soc. Biol.* **57**, 53–66 (2011).
45. Salleh, M. R. Life event, stress and illness. *Malays. J. Med. Sci.* **15**, 9–18 (2008).
46. Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P. & Jullagate, S. Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatr. Dis. Treat.* **10**, 2097–2103 (2014).
47. Moussavi, S. et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858 (2007).
48. Topic, R. et al. Somatic comorbidity, metabolic syndrome, cardiovascular risk, and CRP in patients with recurrent depressive disorders. *Croat. Med. J.* **54**, 453–459 (2013).
49. Lloyd, C. E., Roy, T., Nouwen, A. & Chauhan, A. M. Epidemiology of depression in diabetes: international and cross-cultural issues. *J. Affect Disord.* **142**(Suppl), S22–S29 (2012).
50. Ohayon, M. M. & Schatzberg, A. F. Using chronic pain to predict depressive morbidity in the general population. *Arch. Gen. Psychiatry* **60**, 39–47 (2003).
51. Slavich, G. M. & Irwin, M. R. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol. Bull.* **140**, 774–815 (2014).
52. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
53. Gunderson, K. L. Whole-genome genotyping on bead arrays. in *DNA Microarrays for Biomedical Research: Methods and Protocols* (ed. Dufva, M.) 197-213 (Humana Press, Totowa, NJ, 2009).
54. Kerr, S. M. et al. Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *Bmc. Med. Genet.* **14**, 38 (2013).
55. Nagy, R. et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* **9**, 23 (2017).
56. Navrady, L. B. et al. Cohort profile: stratifying resilience and depression longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *Int. J. Epidemiol.* **47**,13-14g (2018).
57. Smith, B. H. et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *Bmc. Med. Genet.* **7**, 74 (2006).
58. Fernandez-Pujals, A. M. et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
59. Huang, J. et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
60. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS. Med.* **12**, e1001779 (2015).
61. Protocol for a large-scale prospective epidemiological resource. *UK Biobank (2006) www.ukbiobank.ac.uk/resources/.* www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf.
62. UK Biobank Ethics and Governance Framework, Version 3.0. *UK Biobank (2007).* Ethics and Governance Framework. www.ukbiobank.ac.uk/resources/. www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf
63. Brugha, T. S. & Cragg, D. The list of threatening experiences: the reliability and validity of a brief life events questionnaire. *Acta Psychiatr. Scand.* **82**, 77–81 (1990).
64. Motrico, E. et al. Psychometric properties of the list of threatening experiences–LTE and its association with psychosocial factors and mental disorders according to different scoring methods. *J. Affect Disord.* **150**, 931–940 (2013).
65. Goldberg, D. P. & Hillier, V. F. A scaled version of the general health questionnaire. *Psychol. Med.* **9**, 139–145 (1979).
66. Sterling, M. General health questionnaire − 28 (GHQ-28). *J. Physiother.* **57**, 259 (2011).
67. Goldberg, D. P. et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol. Med.* **27**, 191–197 (1997).
68. Wang, L. et al. [Value of patient health questionnaires (PHQ)−9 and PHQ-2 for screening depression disorders in cardiovascular outpatients]. *Zhonghua Xin Xue Guan Bing. Za Zhi* **43**, 428−−431 (2015).
69. Spitzer, R. L., Kroenke, K. & Williams, J. B. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA* **282**, 1737–1744 (1999).
70. Smith, D. J. et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362 (2013).
71. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
72. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
73. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
74. Almli, L. M. et al. Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392–1399 (2014).
75. Keller, M. C. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24 (2014).
76. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: functional mapping and annotation of genetic associations. *Nat. Commun.* **8**, 1826 (2017).
77. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
78. Cross-Disorder Group of the Psychiatric Genomics et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
79. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Commun.* (in press).
80. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *BMJ* **332**, 1080 (2006).
81. Stein, J. L. et al. Voxelwise genome-wide association study (vGWAS). *Neuroimage* **53**, 1160–1174 (2010).
82. Li, Z. et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).
83. Hart, A. B. et al. Genome-wide association study of d-amphetamine response in healthy volunteers identifies putative associations, including cadherin 13 (CDH13). *PLoS One* **7**, e42646 (2012).
84. Sugimoto, K. et al. The induction of H3K9 methylation by PIWIL4 at the p16Ink4a locus. *Biochem. Biophys. Res. Commun.* **359**, 497–502 (2007).
85. Sivagurunathan, S., Arunachalam, J. P. & Chidambaram, S. PIWI-like protein, HIWI2 is aberrantly expressed in retinoblastoma cells and affects cell-cycle potentially through OTX2. *Cell. Mol. Biol. Lett.* **22**, 17 (2017).

86. Lee, H. H. C. et al. Genetic Otx2 mis-localization delays critical period plasticity across brain regions. *Mol. Psychiatry* **22**, 680–688 (2017).
87. Sonuga-Barke, E. J. et al. Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B**, 1359–1368 (2008).
88. Belzeaux, R. et al. Responder and nonresponder patients exhibit different peripheral transcriptional signatures during major depressive episode. *Transl. Psychiatry* **2**, e185 (2012).
89. Mamdani, F., Berlim, M. T., Beaulieu, M. M. & Turecki, G. Pharmacogenomic predictors of citalopram treatment outcome in major depressive disorder. *World J. Biol. Psychiatry* **15**, 135–144 (2014).
90. Wong, M. L. et al. The PHF21B gene is associated with major depression and modulates the stress response. *Mol. Psychiatry* **22**, 1015–1025 (2017).
91. Walsh, R. M. et al. Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. *Nat. Commun.* **8**, 15142 (2017).
92. Shi, G. et al. PHD finger protein 2 (PHF2) represses ribosomal RNA gene transcription by antagonizing PHF finger protein 8 (PHF8) and recruiting methyltransferase SUV39H1. *J. Biol. Chem.* **289**, 29691–29700 (2014).
93. Yu, D. et al. Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *Am. J. Psychiatry* **172**, 82–93 (2015).
94. Eaves, L. J., Last, K., Martin, N. G. & Jinks, J. L. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Br. J. Math. Stat. Psychol.* **30**, 1–42 (1977).
95. Polderman, T. J. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
96. Ceccato, S., Kudielka, B. M. & Schwieren, C. Increased risk taking in relation to chronic stress in adults. *Front. Psychol.* **6**, 2036 (2015).
97. Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A. & Spinath, F. M. Life events as environmental states and genetic traits and the role of personality: a longitudinal twin study. *Behav. Genet.* **42**, 57–72 (2012).
98. Conway, C. C., Hammen, C., Brennan, P. A., Lind, P. A. & Najman, J. M. Interaction of chronic stress with serotonin transporter and catechol-O-methyltransferase polymorphisms in predicting youth depression. *Depress Anxiety* **27**, 737–745 (2010).
99. Cicchetti, D. & Rogosch, F. A. Genetic moderation of child maltreatment effects on depression and internalizing symptoms by serotonin transporter linked polymorphic region (5-HTTLPR), brain-derived neurotrophic factor (BDNF), norepinephrine transporter (NET), and corticotropin releasing hormone receptor 1 (CRHR1) genes in African American children. *Dev. Psychopathol.* **26**, 1219–1239 (2014).